Title page (with author name and affiliations)

A Machine Learning approach for regional geochemical data: Platinum-Group Element geochemistry *vs* geodynamic settings of the North Atlantic Igneous Province

Jordan J. Lindsay[a†], Hannah S. R. Hughes[a], Christopher M. Yeomans[a], Jens C. Ø. Andersen[a], Iain McDonald[b]

[a] Camborne School of Mines, University of Exeter, Penryn Campus, Penryn, Cornwall, TR10 9FE, United Kingdom

[b] School of Earth and Ocean Sciences, Main College, Cardiff University, Park Place, Cardiff, CF10 3AT, United Kingdom

† Corresponding author: e-mail, JL731@exeter.ac.uk

Abstract

Whilst traditional approaches to geochemistry provide valuable insights into magmatic processes such as melting and element fractionation, by considering entire regional data sets on an objective basis using machine learning algorithms (MLA), we can highlight new facets within the broader data structure and significantly enhance previous geochemical interpretations. The platinum-group element (PGE) budget of lavas in the North Atlantic Igneous Province (NAIP) have been shown to vary systematically according to age, geographic location and geodynamic environment. Given the large multi-element geochemical data set available for the region, MLA was employed to explore the magmatic controls on these shifting concentrations. The key advantage of using machine learning in analysis is its ability to cluster samples across multi-dimensional (i.e., multi-element) space. The NAIP data set is manipulated using Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbour Embedding (t-SNE) techniques to increase separability in the data alongside clustering using the *k*-means MLA. The new multi-element classification is compared to the original geographic classification to assess the performance of both approaches. The workflow provides a means for creating an objective high-dimensional investigation on a geochemical data set and particularly enhances the identification of metallogenic anomalies across the region. The techniques used highlight three distinct multi-element end-members which successfully capture the variability of the majority of elements included as input variables. These end-members are seen to fluctuate in prominence throughout the NAIP, which we propose reflects the changing geodynamic environment and melting source. Crucially, the variability of Pt and Pd are not reflected in MLA-based clustering trends, suggesting that they vary independently through controls not readily demonstrated by the NAIP major or trace element data structure (i.e., other proxies for magmatic differentiation). This data science approach thus highlights that PGE (here signalled by Pt/Pd ratio) may be used to identify otherwise localised or cryptic geochemical inputs from the subcontinental lithospheric mantle (SCLM) during the ascent of plume-derived magma, and thereby impact upon the resulting metallogenic basket.

Keywords: *Platinum-group elements; machine learning; plume; geochemistry; metallogenesis; mantle*

Abbreviations: *platinum-group element(s) (PGE); North Atlantic Igneous Province (NAIP); Principal Component Analysis (PCA); t-stochastic neighbour embedding (t-SNE); subcontinental lithospheric mantle (SCLM); machine learning algorithm(s) (MLA); base metal sulphide(s) (BMS)*

1. Introduction

1.1 Data science in geochemistry

Machine learning is a powerful data science tool used to investigate large datasets and is increasingly integrated into novel scientific applications. Bulk geochemical data sets are excellent targets to analyse using machine learning algorithms (MLAs) considering they often comprise large sample sets with a multitude of elements measured for each sample. Furthermore, global and regional geochemical data sets are becoming less expensive to create, and more accessible through analytical development and data sharing capabilities. Previous studies have successfully implemented MLA-based methodologies to investigate geochemical domains on Mars (Taylor et al., 2010), hazardous groundwater geochemistry (Farnham et al., 2002), and Pb behaviour in the Kerman Copper Belt, Iran (Ghannadpour et al., 2013). Recently, efforts have been made to enhance the link between traditional geochemical investigations and data science, such as the revisiting of traditional basalt geochemical sub-groups (Iwamori et al., 2017) and mapping mineral distributions in lunar basalts (Cone et al., 2020) using multivariate statistical analyses.

Herein, this paper tests a combined classical and MLA approach applied to a large magmatic geochemical data set, with the aim to establish a framework that can be replicated for a variety of studies in the field of geochemistry. The comparison of descriptor-based labelling (e.g., geographic location or lithological classification) and algorithmic clusters is of particular interest in geochemistry, since geochemical studies often partition data based on a small number of variables selected by the user and do not incorporate the larger-scale similarities across all variables. In this respect, simplifying data sets may resolve major trends, but bypass more subtle relationships that exist across multiple elements and that could enhance the interpretation. Machine learning provides a means for more sophisticated analysis alongside classical geochemical techniques.

By exploring an example data set, the bulk geochemistry of North Atlantic Igneous Province (NAIP) lavas, we contribute to the discussion surrounding mantle plume and subcontinental lithospheric

mantle (SCLM) controls on platinum-group element (PGE) and precious metal abundances in basaltic magmas. The application of data science techniques such as MLAs allows for the critical examination of elemental concentrations from a different, objective perspective by analysing geochemistry in multi-dimensional space – something not attainable through traditional geochemical data analyses. The workflow presented in this paper, including dimensionality reduction methods and clustering MLA, can be used for a variety of similar studies alongside prior or concurrent classical discriminant diagrams, for example, forming a framework for comparisons of mantle plume geochemical signatures at a global scale.

1.2 Geological background

1.2.1 Mantle plumes and continent break-up

Mantle plumes are hot upwellings from the lowermost portions of the silicate Earth. Initiated by chemical or physical instability at depth (Kellogg and King, 1993; Bercovici and Kelly, 1997; Jellinek and Manga, 2004), they form rising diapirs of buoyant high-temperature mantle material that can induce decompression-driven partial melting of the upper mantle, asthenosphere and lithosphere (Griffiths and Campbell, 1990). Continental landmasses uplifted and thinned by impinging plumes will eventually extend and rift apart, ultimately leading to the formation of oceanic lithosphere (Pirajno and Santosh, 2015). The material melted by plumes under continental crust changes through their lifetime as the geodynamic environment shifts towards an oceanic setting, and the geochemistry of subsequent lavas produced from plume magmas will reflect these changes (Howarth and Harris, 2017). Decompression models for flood basalt melting predict < 5% of the lithosphere is involved in melt generation, and this is isolated to the initiation stages (i.e., when the continental lithosphere is available to melt prior to rifting; McKenzie and White, 1989; White and McKenzie, 1995). If the lithosphere is thick, larger amounts of crust and the SCLM are involved in melt generation; underneath lithosphere thinned by extension, melts will incorporate higher

proportions of asthenospheric (i.e. plume-derived) material (Turner et al., 1996). In addition to the reduction of continental contamination of plume magmas with time, initial melts coinciding with the buoyant, laterally expansive and voluminous plume head are typically interpreted to have stemmed from higher degrees of partial melting and produced extensive eruption of flood basalts (Campbell and Griffiths, 1990). With time, the degree of partial melting may decrease as the narrower plume tail becomes the primary melting source (e.g., Griffiths & Campbell, 1990; Trela et al., 2015).

1.2.2 The North Atlantic Igneous Province

The Icelandic hotspot, the surface expression of the underlying (proto-)Icelandic mantle plume, first erupted lava c. 65 Ma (Berggren et al., 1995), heralding the opening of the North Atlantic Ocean in the Palaeogene (Hole and Natland, 2019). Greenland and the British Isles began to rift apart while mantle plume-derived magmas fed volcanism persisting into the newly opened ocean basin (Saunders et al., 1997; Kent and Fitton, 2000). Collectively, the igneous rocks produced from this event belong to the NAIP (Horni et al., 2017) – Figure 1a. Today, intraplate and rift volcanism are active simultaneously on Iceland (McKenzie and White, 1989; White and McKenzie, 1995; Momme et al., 2003), and the plume has transitioned through continental to oceanic geodynamic settings throughout the last ~ 62 Myr (Fig 1b).

The Icelandic hotspot extruded (and continues to extrude) unusually high volumes of lava with respect to similar plumes around the world (Courtillot et al., 2003), especially in its earlier stages (e.g., Greenland and the British Palaeogene Igneous Province or BPIP) (Larsen and Pedersen, 2000). Lavas in the region are primarily basaltic but range from picrites to more evolved andesites and rhyolites (Kent and Fitton, 2000). The region is recognised as having high potential for hosting Ni-Cu-PGE deposits, partly attributed to its plume setting (Andersen et al., 2002) and has been a focus for PGE research in the last two decades. In some PGE-mineralised localities elsewhere in the world, such as the Bushveld (South Africa) and Stillwater (North America) complexes the suggestion that

104 the SCLM plays a key role in contaminating plume magmas with the metals that form ore deposits
105 through mineralisation remains controversial, yet persistent (e.g., Maier and Groves, 2011 and
106 references therein). Intrusive complexes in the NAIP, such as Skaergaard in East Greenland and the
107 Rum Layered Complex in western Scotland, host mineralised PGE reefs and a key debate concerns
108 the source of these metals within the NAIP magmatic system (e.g., Andersen et al., 1998, 2002;
109 Butcher et al., 1999; Pirrie et al., 2000; Hughes et al., 2015, 2017).
110 In one study, Hughes et al. (2015) demonstrated a systematic shift in the relative PGE abundances of
111 NAIP basaltic lavas from oldest to youngest (e.g., continental to oceanic). Lavas in West Greenland
112 and the BPIP (the earliest products of plume magmatism) typically have Pt/Pd ratios of 1.9; later
113 lavas in East Greenland and its offshore regions have Pt/Pd ~ 0.79; and contemporary Icelandic lavas
114 have Pt/Pd ~ 0.4 (Fig. 1a). The metal signature of NAIP plume-derived melts (i.e., the metal basket)
115 appears to have changed alongside the chemo-dynamic setting. After critiquing a variety of possible
116 explanations for this Pt/Pd shift, Hughes et al. (2015) suggested the contaminating influence of
117 SCLM-derived melts and metals on the plume-derived basaltic magmas was the most likely control.
118 More specifically, mantle peridotite xenoliths entrained within Scottish lamprophyre dykes,
119 representative of the mineralogy of the SCLM underneath the current margins of the North Atlantic
120 craton, contain two populations of base metal sulphides, one of which is notably Pt-rich due to their
121 inclusion of PtS, the platinum-group mineral cooperite (Hughes et al., 2017). If plume magmas from
122 the earlier stages on the NAIP (e.g., West Greenland and Scotland) incorporated portions of the
123 SCLM during asthenospheric melting, assimilation of such platinum-group minerals (PGM) would be
124 reflected in higher Pt/Pd in bulk geochemistry of the lavas produced. This effect would reduce in line
125 with the progressively decreasing role of SCLM contamination of plume-derived magmas in oceanic
126 settings (e.g., more recent East Greenland offshore and Iceland lavas), reducing Pt/Pd ratio and
127 therefore signalling an inherent linkage between enriched mantle keels and metal prospectivity (e.g.,
128 Hawkesworth and Scherstén, 2007; Hughes et al., 2014).

In this paper, we seek to test the validity of the apparent shift in Pt/Pd ratio by assessing the precious metal compositions of NAIP lavas in context with their major and trace element geochemistry across multi-dimensional (multi-element) space. By using ML to objectively classify the high-dimensional trends in the data, we seek to comment on the major processes captured in a regional-scale magmatic differentiation system and whether controls on metallogenesis are truly localised according to geodynamic setting.

2. Methods

The purpose of using machine learning to analyse the NAIP data set is to explore different facets and enhance our understanding of a previously-studied geochemical system. Geochemistry has, until recently, rarely been coupled with data science (Zuo, 2017) and there are significant opportunities to develop an integrated geochemical workflow for high-dimensional data analysis using MLAs, as proposed in Figure 2. Traditional geochemical investigations are essential for understanding the geological processes behind magma compositions, but a data science approach is hereby used to maximise the information obtained from such data sets and complement findings from earlier investigations. Dimensionality reduction techniques like Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbour Embedding (t-SNE) describe the large-scale variability of a large data set in an intuitive way. A variety of MLAs process large amounts of data and identify high-dimensional trends unresolvable to a human analyst.

The ratio of Pt/Pd changes across NAIP lavas is documented by Hughes et al. (2015), as reflected in their changing geography. By using a data science approach supported by dimensionality reduction, we investigate how elemental concentrations behave with respect to each other and geographic information. The large-scale structure of the data will inform the investigation of controls on shifting PGE ratios (and other elemental concentrations) while directly comparing the new findings to those from earlier studies.

154

155 ***2.1. Data***

156 The NAIP data set from Hughes et al. (2015) contains new bulk concentrations of 49 major and trace
157 elements from Scottish basalts in addition to existing literature data for basaltic lavas from West and
158 East Greenland, offshore Greenland, and Iceland (Table 1; Figure 1). Hughes et al. (2015) used their
159 own data alongside a compilation of other PGE-bearing data sets from across the NAIP for their
160 study (Table 1) – the amalgamated data set is presented in Supplementary Data A. The data were
161 classified by locality (i.e. five categories) and the study compared major, trace and PGE
162 concentrations in lavas between these locations. Our study uses new methods to produce different
163 categories based on multi-element geochemical variability, to determine how well sample localities
164 (and by extension, geodynamic stages in the NAIP) are reflected in numerical clustering techniques.

165 The dimensionality reduction and machine learning techniques used in this study necessitate a
166 complete data set with no missing information. The number of elements analysed and available in
167 each of the five sample sets varied greatly, leaving eleven variables present in every North Atlantic
168 locality with mainly non-zero concentrations – major element oxides Fe_2O_3, MgO and TiO_2, and trace
169 elements Cr, Ni, Cu, Ir, Ru, Rh, Pt, and Pd. This is by no means an exhaustive variable set, missing
170 lithophiles, alkalis and rare earth elements, but by itself can provide a great deal of information
171 towards large-scale plume melting processes. The eleven modelled elements are important proxies
172 for magmatic differentiation processes in mafic magmatic systems and are key for mineralisation of
173 Ni-Cu-PGE ores, explaining their consistent inclusion in all five sample sets.

174 Of these selected variables, where occasional blank/non-numerical observations existed, a process
175 of rounded-zero imputation was applied following the approach by Martín-Fernández et al (2012),
176 using the mean least-squares regression values against all other variables. This gives previously
177 missing observation cells meaningful concentrations below the elemental detection limits, without
178 implying detectable concentrations or disrupting the variance of the data as a whole. The final data

179 set used is supplied in Supplementary Data A. The following sections describe the workflow used to
180 analyse the data.

181 To avoid biasing the clustering through variables with mixed-unit concentrations (major oxides in
182 percent, minors in ppm and PGE in ppb) z-scores were generated for the data set as a
183 standardisation procedure. The following equation was used to produce z-scores for all data
184 observations (as per Kreyszig, 1979) when z is the z-score, x is the raw concentration, μ is the
185 population mean and σ is the standard deviation of the population:

186 $$z = \frac{x - \mu}{\sigma}$$

187 Standardising the data ensures that variability in a given variable can be directly compared to
188 another, regardless of raw data units, leading to more effective clustering.

189

190 ***2.2 Principal Component Analysis***

191 Popularised by Hotelling (1933) after Pearson (1901), PCA is a multivariate analysis tool commonly
192 used to transform a data set in such a way that the variances of each dependent variable can be
193 viewed in unison in a low dimensional space (Davis, 2002; Jolliffe, 2002 and references therein). It
194 functions as an effective linear dimensionality reduction technique. The method creates individual
195 Principal Components (PCs), which describe the contributions to data variance, where the first order
196 PC is aligned to capture the maximum variance in the data spread, the second order PC will then be
197 aligned orthogonal to the first and so on. The main purpose of using PCA for the NAIP geochemical
198 data set is to determine how elemental concentrations in the basaltic lavas are correlated to these
199 PCs and by extension, which concentrations are controlled by similar factors. PCA also allows us to
200 display multi-element information in biplots (Hyvärinen et al., 2001) and effectively discuss the
201 structure of the data set using minimal dimensions (Chang, 1983). For a large geochemical data set,
202 PCA can describe the contribution from each element to the overall variance of the data set,

allowing for a simple comparison of correlations between elemental concentrations, analogous to the relationships for the eleven NAIP elements shown in a correlation matrix (Fig. 3). Elements with high correlation coefficients will likely be represented in the same PCs, and thus, similar reflect underlying controls.

2.3 *t-SNE*

Similar to PCA, t-SNE is a dimensionality reduction technique used to display multi-dimensional data in an easily interpretable manner. The algorithm was developed by van der Maaten & Hinton (2008) and translates high-dimensional data (with each dimension representing one variable from the data set) into a low-dimensional bivariate space using two new computed features referred to as *embeddings*. Using the Kullback-Leibler Divergence (Kullback and Leibler, 1959), the algorithm maximises the similarity between the positions of all data points in high-dimensional space and their position in the embedding plot. The advantage of this approach is that it retains the original data structure while significantly reducing dimensionality, compared to PCA. Data points with similar multi-element concentrations will plot closely in the resulting two-dimensional embedding (van der Maaten and Hinton, 2008; van der Maaten, 2014). This technique acts as a data structure map and the transformed data points can be overlain with their MLA-based or descriptor-based classifications. Contributing element concentrations can also be mapped on to the embedding space, to determine how their abundances are distributed through the overall data structure..

Studies have successfully used t-SNE as a geochemical discriminant tool for alteration indicators (Horrocks et al., 2019) and geological domain mapping (Balamurali and Melkumyan, 2016), and the method is applied in a similar fashion to the NAIP data to identify metallogenic signatures. Here, we use the t-SNE algorithm from the *sci-kit learn* package in Python 3.7.4 (Pedregosa et al., 2011) to resolve and display the z-scored elemental contributions to overall NAIP data structure, with close reference to PCA results derived herein. Key input parameters were set up is as follows: perplexity of

80, 5000 maximum iterations and learning rate of 200, based on recommendations for 259 samples by Krijth (2015). Geographic categories from Hughes et al. (2015) and MLA-based clusters developed in this study are overlain on the generated t-SNE embedding, to allow for a simple comparison of the different clustering methods.

2.4 *k*-means Clustering

Machine learning techniques fall into three major categories: supervised learning, in which you train an algorithm on a portion of a data set to analyse or make predictions about the remaining portion; unsupervised learning, in which the algorithm finds its own structure in the data without the need for class labels; and reinforcement learning, where algorithms perform sequences of decisions based on both exploration and exploitation of knowledge (e.g., Hastie et al., 2009; Marsland, 2009; Witten et al., 2016). One of many unsupervised machine learning algorithms used to cluster multivariate data is the *k*-means clustering technique (MacQueen, 1967) which has been selected as an analytical technique for the NAIP data set on account of its versatility, ease of use and ability to cluster data based on user-selected parameters. The distance-based algorithm partitions a data set based on similarities amongst a large number of variables so that occurrences within the same grouping are more similar to each other than occurrences in another grouping (e.g., Michie et al., 1994; Hastie *et al.*, 2009; Marsland, 2009). The analyst must select the desired number of clusters (*k*) and the algorithm then randomly assigns *k* centroids to the data set. All observations are attributed to the nearest centroid to form clusters. The algorithm repositions the centroids and iterates until the sum of square Euclidian distances from each data point to their mean centre are minimised across the entire data set, to find the optimum centroid positions and corresponding clustering formation (Howarth, 1983). Crucially, this technique can be used with a high-dimensional data set, something unattainable via manual interpretations, and consequently will produce more objective and

252 statistically-robust clusters than a manual clustering exercise. For the chosen set of variables,
253 screening of outliers was unnecessary.

254 The unsupervised algorithm does not require *a priori* sample labels to produce clusters and as such,
255 the optimum *k* can be difficult to discern; a heuristic approach is recommended when selecting *k*-
256 values, which promotes running the model multiple times with different input parameters to achieve
257 a satisfactory result. The Davies-Bouldin Index (DBI) (Davies and Bouldin, 1979) can be used to
258 retrospectively assess the statistical performance of different input parameter setups in a model, but
259 model selection should also consider more subjective qualifiers in the context of the data set.
260 Ideally, classifications should have small intra-cluster distances and high inter-cluster distances. In
261 this study, the *k*-means algorithm from the sci-kit learn package in Python 3.7.4 (Pedregosa et al.,
262 2011) was implemented including the z-scores of (i) all 11 available variables, (ii) PGE and trace
263 elements Cr, Ni and Cu, (iii) PGE only, and (iv) a selection of generated PCs (non-z-score), to observe
264 the differences these inputs had on cluster size, shape, placement and DBI.

265

266 ## 3. Results

267 ### *3.1 Dimensionality reduction*

268 #### *3.1.1 Principal Components*

269 PCA was performed for the eleven elements in the NAIP data set. PCs 1 through 6 account for
270 92.17% of the variability in the NAIP data set combined (43.99%, 20.06%, 13.53%, 6.78%, 5.11% and
271 2.70%, respectively). A scree plot is provided in Figure 4a that summarises the relative importance
272 of each PC. As a rough guide, eigenvalues beyond ~90% cumulative contribution are normally
273 considered to be background noise and not substantially adding to data set variability, denoted by
274 the flattening of the scree slope. PC7 and onwards are superfluous and not discussed herein. Full
275 PCA statistics, including eigenvalues, eigenvectors and scaled co-ordinates are displayed in
276 Supplementary Data B. Figure 4b-d displays PC score biplots (combined variable and sample

information) for (b) PC1-PC2, (c) PC3-PC4, and (d) PC5-PC6. Eigenvectors for each element are plotted as lines from the origin, and individual samples are plotted and attributed to their geographic categories (Hughes et al., 2015). Biplots allow for simultaneous interpretation of both element and sample variance. Vector lengths and directions represent the degree to which each PC describes the variability of the corresponding element.

As shown in Figure 4b, there are three major element vector groups in the PC1-PC2 space: MgO, Ni, Cr, Ir and Ru, positively attributed to PC1; Fe_2O_3, TiO_2 and Cu, positively attributed to PC2; and Pd, Pt and Rh, positively attributed to both PC1 and PC2. In Figure 3, these groups tend to have positive correlation coefficients with other members of their group. In Figure 4c, Pt, Pd and Rh retain their association from Figure 4b and are the only variables to plot strongly positively against PC3; all other elements plot to the left hand side of the PC3-PC4 space, with the MgO-Ni-Cr-Ru-Ir group displaying shorter vector lengths than the Fe_2O_3-TiO_2-Cu group. The majority of the variables have negligible to negative association with PC4, with only Pt, Ir and Fe_2O_3 vectors plotting positively. In Figure 4d, Pt and Pd appear correlated with the previously identified MgO-Ni-Cr group in the negative direction of both PC5 and PC6. In this space, TiO_2 exhibits an opposite vector direction to Fe_2O_3 and Cu. Finally, Ir, Ru and Rh exhibit positive trends with PC5, with increasing vector length in that order; Ir and Ru are positively correlated with PC6. In terms of sample variance, onshore West and East Greenlandic lavas are coupled with MgO and its correlated vectors discussed above, while East Greenlandic (offshore), Icelandic and BPIP lavas plot opposed to most multi-elemental vector groups in higher order PCs (Fig. 4b and 4c). It appears that in some instances (Fig. 4c and d), extreme West Greenlandic outliers with high single-element contributions (notably Pt, TiO_2 and Rh) have an influence on eigenvector length.

3.1.2 Embedding

A t-SNE analysis of the NAIP data set created using the z-scores of all eleven elements as input variables is displayed in Figure 5. The transformed data distribution is coloured by the individual elemental concentrations for each data point, giving an account of how each element contributes to the variability of the entire data set. The newly computed features, Embedding 1 and 2, arrange the data points in a roughly spherical shape with prominent protrusions on the top-left, top-right and bottom-middle of the distribution. Elements that exhibit clear bimodality through the embedding significantly contribute to the data set structure and corresponding low-dimensional shape (Horrocks et al., 2019). For example, high concentrations of MgO are found in the bottom of the embedding, and data points outside this zone have reduced MgO content comparatively. From this logic, the inverse relationship between MgO-Ni-Cr-Ir-Ru and TiO_2-Fe_2O_3-Cu-Pd trends from PCA (Fig. 4b-d) is clear along the length of Embedding 2 in the form of opposed bimodal distributions. A smaller separation is also visible between Cu (top-left), TiO_2 (top-right) and Fe_2O_3 (top-spread) the uppermost zone of the embedding. A third group of Pd, Pt and Rh express bimodality along Embedding 1, particularly the former, with all three plotting highest values on the left of the embedding. This is in direct opposition to the major bimodality distribution shown in the other elements. Extreme outlier values of Pt and Rh concentrations do not interfere with this trend. The ratio of Pt/Pd does not appear to have any distinctive trends in the embedding. As per their correlation in Figure 3, Pd and Cu are seen to have reasonable overlap in the embedding. In addition to overall trends, t-SNE can isolate data set anomalies within the context of the data structure and attribute their segregated nature to single or multi-element concentrations. Sub-clusters can be seen to host distinctive concentrations pertaining to either the MgO or TiO_2-led trends from Figure 4, and high single element concentrations define individual outlier samples e.g., the segregated high-Ni (top-centre), high-Rh (bottom-right) and Pt (bottom-left) occurrences in the embedding.

3.2 MLA model selection

The NAIP data set was clustered using the *k*-means algorithm for *k*-values of 2, 3, 4 and 5, with the different variable arrays detailed in the methodology, including variable z-scores and PC1-6 (selected as per Fig. 4a). Davies-Bouldin Indices, which account of cluster performance of all configurations, are shown in Figure 6. A high Davies-Bouldin Index (DBI) signifies that the density function of each defined cluster is larger, meaning more diffuse or looser clustering. In this instance, a low DBI is desirable, reflecting tighter, more well-defined clusters (Davies and Bouldin, 1979).

Models using *k*-values of 2, whilst consistently having some of the lowest Davies-Bouldin Indices of the model set (< 1.40), appear too simplistic to describe the evolving geochemistry of the NAIP lavas, especially considering the number of groups defined by Hughes et al. (2015) were either 3 (Pt/Pd subdivisions) or 5 (geographic categories). After discounting *k*=2 models due to their simplicity, and all models using z-scores of raw concentrations as variables due to their higher Davies-Bouldin Indices, the parameter setup of *k*=3 with PC1-6 as input variables is selected as the best model (highlighted by the black square in Fig. 6). It is clear that in data sets with large variable numbers, using new features produced via dimensionality reduction is vital to help focus high-dimensional trends, recognise multivariate structures and create robust clustering models (Nguyen and Holmes, 2019). Figure 7 shows a comparison of clustering models using PC1-6 as variables and *k* ranging from 2 to 5. This is useful to demonstrate how high-dimensional clusters from different parameter set-ups relate in a bivariate space, in the case of Figure 7, PC1 against PC2. Surplus models are displayed in Supplementary Data C and all clustering results compiled in Supplementary Data D.

3.3 Clustering

The clusters created by the chosen *k*-means algorithm are referred to as Group 1 (red circles), Group 2 (blue squares) and Group 3 (yellow diamonds). It is convenient to display newly assigned multivariate cluster information in bivariate plots to determine the relative contributions each element's variability made to the overall cluster formation. Elements that are important to overall

data structure will have easily identifiable and distinct clusters along the corresponding axes of a bivariate plot. Figure 8 displays algorithm-clustered bivariate plots of elemental concentrations from the NAIP data set. Major element oxides MgO and TiO_2 cluster with a progressive negative trend (Fig. 8a). By contrast, Fe_2O_3 does not exhibit a linear trend, with Groups 1 and 2 perhaps exhibiting slightly higher concentrations in a wider range than Group 3 (Fig. 8b).

Trace elements Cr and Ni, and Ir, Ru and Rh from the PGE all cluster neatly in Figures 4b-f, with the highest concentrations of each belonging to Group 3 (corresponding to higher MgO and lower TiO_2); this again reflects their high correlation coefficients with each other in Figure 3. Copper does not cluster as distinctly as the other trace elements, but a slight negative relationship with Ni is visible, with the highest Cu concentrations in Group 1 and comparable low concentrations in Groups 2 and 3 (Fig. 8c). The correlation matrix for the NAIP data set (Fig. 3) accordingly describes Cu as lacking in strong affinity with any other elements other than TiO_2 and Pd. It appears that Group 2, the smallest by sample number, exists as an intermediate cluster with no significant enrichments, whereas Groups 1 and 3 have enrichments in elements from opposing PCA and t-SNE trends (Figs. 4b-d and 5) and act as geochemical end-members in the system. Overall, Group 1 is defined by higher concentrations of TiO_2, Fe_2O_3 and Cu, Group 2 is defined by background to very low concentrations for most elements and Group 3 is defined by higher concentrations of MgO, Ni, Cr, Ir Ru and Rh.

There are no distinct clustering trends in two particular variables – Pt and Pd. Figure 9a shows an MLA-clustered bivariate plot of Pt and Pd, and highlights the lack of definition in both variables. Group 2 appears to host the lowest concentrations for both elements, although all three groups share similar distribution of higher concentrations and clusters are indistinct. Figure 9b illustrates a similar absence of trend in terms of Pt/Pd ratio as a function of Pd.

Overall, the *k*-means model successfully clustered data in nine elements based on the chosen input parameters of *k*=3 and PC1-6 as variables. The remaining two elements, Pt and Pd, present an interesting anomaly in the clustering process, and this would appear to be in agreement with the

lack of strong correlation coefficients (Fig. 3), unique PCA characters (Fig. 4) and t-SNE concentration trends opposed to all other elements (Fig. 5) for these two elements, with a weak to moderate relationship with Rh in some cases.

3.4 Comparison to prior NAIP categories

A key check on the effectiveness of using the MLA-integrated workflow as a geochemical tool is comparing algorithm-based clustering to information already established by classical studies using discriminant plots, to identify similarities and differences in element behaviour. Figure 10 displays a histogram of the five geographic groups and their new equivalent cluster distribution based on the chosen *k*-means model. The Iceland group comprises similar proportions of Group 1, 2 and 3 samples (roughly a third of each). BPIP lavas are 51% within Group 1, 24% within Group 2 and 25% within Group 3, making a roughly 2:1:1 split. East Greenland offshore and onshore lavas belong largely to the Group 1 end-member (71% and 75%, respectively); they differ in that the former contains small proportions of Group 2 and 3 and the latter is completely devoid of Group 2 in the chosen clustering set-up. Finally, West Greenland contains the largest proportion of Group 3 samples (48%) and the smallest proportion of Group 1 samples (29%) in the region.

A further effective visual comparison of the MLA workflow results can be achieved by overlaying data distributions in bivariate plots and t-SNE embeddings with new clusters and geographic categories. A selection of key bivariate PGE relationships is shown in Figure 11. Trends and anomalies in the newly clustered classifications both complement and differ from the NAIP geographic categories in a variety of manners.

As shown in Figure 11a, the average Pt/Pd ratios neatly described by geographic groupings (Figs. 1a and 11b) by Hughes et al. (2015) do not form as clear a trend with the *k*-means clustering method, which echoes the lack of distinct clusters for those individual variables in Figure 9. A very slight

400 decrease in average Pt/Pd is observed from Group 3 to 2 to 1, but this is on a much smaller degree
401 than the trends in geographic categories. It should be noted that although the regression lines for
402 Pt/Pd in East Greenland (onshore) lavas were not as distinct as other localities in the original study
403 (Momme et al., 2002; Hughes et al., 2015), the basic age-progressive reduction in average Pt/Pd
404 does persist throughout, using the data available.

405 In Figure 11c and 11d, plots of Ir vs. Pd/Ir ratio convey trends typical of olivine accumulation and
406 incompatible element fractionation as controls on PGE distribution within magmas. Group 1 plots
407 neatly along the incompatible fractionation trend and Group 3 plots with olivine accumulation.
408 Group 2 does not appear to have a strong affinity with either trend, plotting in low Ir regions far
409 from the major fractionation population (Fig. 11c). In Figure 11d, West Greenlandic lava Pt/Pd ratio
410 is attributed mainly to olivine accumulation (and thus, correlated with Group 3 as per Figure 10).
411 East Greenlandic (onshore) and, to a lesser extent, Icelandic samples correspond well with Group 1
412 being driven by the fractionation of elements incompatible in modal silicate minerals (e.g., Cu and
413 Pd as per Keays & Lightfoot, 2007; Naldrett, 2004). The geochemistry of the East Greenlandic
414 (offshore) lavas, BPIP lavas and a portion of the Icelandic lavas do not correspond well with either
415 control, with a comparable distribution to Group 2 (Fig. 11c) off the main trend axis – these locations
416 also have the highest proportion of points designated to Group 2 in Figure 10.

417 Figure 12 displays the embedding from Figure 5 created using t-SNE, with all data points classified by
418 (a) *k*-means clustering using the selected model and (b) geographic groupings, summarising how the
419 different clusters interact with the broad structure of the data set. As detailed in van der Maaten &
420 Hinton (2008), t-SNE retains important multivariate data information in a bivariate plot. Groups 1, 2
421 and 3 plot in distinct regions within the newly created embedding in Figure 12a, with Group 1
422 distributed mainly in the centre and left, Group 2 in the upper right and Group 3 in the bottom of
423 the distribution. One notable outlier of Group 2 is located in the far right of the embedding. The split
424 between Group 3 and the other two clusters is similar to their relationships in Figure 8, where Group

3 often plots distinctly from all other points. Groups 1 exhibits the largest spread in variability, physically taking up the most space in the embedding. In contrast, Groups 2 and 3 have much more concentrated distributions in the embedding, which likely relates to their more consistent individual element concentrations throughout Figure 8. Group 3 generally acts as the 'anomalous' cluster of highest MgO, Ni, Cr, Ir and Ru, and lowest Fe_2O_3, TiO_2 and Cu concentrations (Fig. 8a-f), and these enrichments and depletions evidently create a distinctive character for each group's chemistry in Figure 12a in strong agreement with individual element trends in Figure 5. Although less distinct, Group 1 shares a sector of the embedding with higher concentrations of Fe_2O_3, TiO_2 and Cu (Fig. 5).

In general, the t-SNE embedding is not clustered as clearly when classified by geographic location in Figure 12b, but structures are still observable. The majority West Greenlandic lavas (bottom-right) have multi-elemental compositions physically opposed to onshore and offshore East Greenlandic lavas (centre and top-left) in the embedding. Icelandic lavas are concentrated in the boundary between West Greenland and East Greenland (offshore) in the centre of the embedding, and the bulk of BPIP samples appear in the top of the embedding. A large proportion of West Greenland lavas and around half of Icelandic lavas plot in the section of the embedding strongly associated with high MgO, Ni, Cr, Ir, Ru and Rh in Figure 5 and Group 3 in Figure 12a (which agrees with large proportions of their lavas belonging to Group 3 in Fig. 10). East Greenland (offshore and onshore) are associated with high Fe_2O_3, TiO_2 and Cu (Fig. 5) and Group 1 (Fig. 12), as per Figure 10 (< 70% Group 1 lavas). BPIP appears evenly spread between prior identified regions in the embedding. Individual sub-clusters forming distinct populations outside and within the main data structure and can be observed in Figure 12a and 12b. These were defined subjectively as breaks from the otherwise continuous cumulative variability map and are always dominated by a particular high concentration in Figure 5, a particular Group in Figure 12a or a particular locality in Figure 12b. Selected prominent sub-clusters are indicated by numbered annotations.

A series of summary box-and-whisker plots with elemental concentrations for each MLA-based cluster is given in Figure 13 to summarise their typical geochemical signatures. Even in these simple plots, the lack of significant inter-cluster Pt and Pd variations are clear when compared to the other elements, which normally see reasonable shifts in quartile ranges between Groups 1, 2 and 3.

4 Discussion

4.1 *Performance of MLA-based geochemical workflow*

Utilising data science techniques to explore the NAIP data set has allowed us to analyse and display data in a variety of ways not attainable by a classic interrogation of geochemistry. Not only does establishing a MLA-integrated workflow save processing time and provide a framework easily applied to other data sets, it reveals unique information about how elemental concentrations vary alongside each other. The correlation matrix (Fig. 3) delineates which elements varied with respect to each other, expanded upon by PCA in Figure 4. At least two multi-elemental geochemical end-member associations exist within the broad data structure – a high MgO group, and a high-TiO_2 group. t-SNE displayed complete chemical variability information about NAIP lavas across eleven elements, while simultaneously identifying anomalous multi-element sub-clusters distinct from the main population and attributed to a particular grouping. The technique further reduced dimensionality and independently replicates the MgO- and TiO_2-led end-members identified by PCA in the context of data structure. In both PCA and t-SNE dimensionality reduction results, Rh, Pt and Pd are highlighted as not strongly conforming to either of the major end-member groups. Sub-clusters identified by t-SNE can be investigated in isolation and are often attributed to individual sample localities; it is likely that these distinct masses represent a unique chemistry still conforming to the overall algorithm-based clusters but with individual elemental concentrations outside the norm. This finding agrees with Figure 5, in which sub-clusters can often be explained by particularly high concentrations of a single variable.

474 Using machine learning to cluster the data set beyond arbitrary single element concentrations
475 further explores these multi-elemental end-members and classifies data points broadly in line with
476 findings from dimensionality reduction. The intricacies of a high-dimensional data set can be
477 resolved and presented in a logical manner, with the three techniques complementing and
478 enhancing the findings from other steps; multi-element patterns can be observed through all three
479 techniques (Figs. 4, 5 and 8), PC features were directly used in the best-fit clustering model (Figs. 6-
480 9), and t-SNE was used to visualise overall differences in MLA-based classifications and geographic
481 categories from other steps in the workflow (Fig. 12). Whilst the NAIP variable set of eleven
482 elements is certainly large enough to merit exploration using a MLA approach and the applicability
483 of dimensionality reduction prior to clustering is reflected in Figure 6, this would become
484 increasingly relevant as a data set integrates more variables. By condensing large-scale elemental
485 behaviour into fewer features whilst retaining variability information, the workflow makes
486 geochemical interrogation significantly more manageable.

487 Despite the success of the workflow results, some caveats must be acknowledged when utilising
488 clustering algorithms and dimensionality reduction for geochemistry. The selection of input
489 parameters for most machine learning techniques is an analyst-dependent pursuit – *k*-means
490 clustering cannot optimise parameters and model selection requires a set of considerations unique
491 to the job the algorithm will be applied to. Furthermore, while clustering data points based on a
492 large set of elements can be viewed as objective by encompassing all variable concentrations,
493 elements that do not conform to the overall variance of the data set can impede perceived
494 algorithm success. Similarly, if certain input parameter set-ups are chosen (e.g., with inefficient *k*-
495 value and t-SNE input parameters, non-normalised data or incomplete variable sets) any of the
496 models used in this study may be unsuccessful or unrepresentative. However, as long as
497 considerations are made regarding the set-up of the methodology, a MLA approach does not
498 become a 'black box' process, and offers new prospects in understanding a data set. While data
499 science provides important and unique information about a data set, a reasonable knowledge of the

subject being investigated is essential to aid in the interpretation of results – a careful balance must be maintained between sensible user input to drive MLA success and background geochemical understanding of the project without bias. The following discussion takes findings from the NAIP workflow – namely the three multi-element clusters and the seemingly isolated behaviour of certain PGE – and contextualises them by relating data anomalies and trends to geochemical processes.

4.2 *MLA mapped on to geochemical indicators for magmatic differentiation in the NAIP*

Further to establishing a new approach for the analysis of regional geochemical data sets, a major feature of this study concerns the behaviour of groups of elements in relation to the data structure. In PCA, t-SNE and *k*-means clustering methods, distinct variability trends are exhibited by most of the elements included in MLA-based analyses. Magnesium (as MgO), Cr, Ni, Ir and Ru are strongly controlled by PC1 (Fig. 4b) and Fe_2O_3, TiO_2 and Cu are strongly controlled by PC2 (Fig. 4c); the two sets of elements also replicate similar trends in t-SNE embeddings in Figure 5 through a different mathematical process. Platinum, Pd and Rh exhibit a mixed influence of PC1-2, and PC3 (and to an extent PC4) further isolates them from every other element. The lower-influence PC5 and PC6 separate the previous elemental groups (Fig. 4d), although their cumulative importance to data variability is much lower (Fig. 4a). These multi-element patterns are broadly reflected in clustering in Figures 7 to 9.

The concentration of an element in a magma reflects the combined effect of several differentiation processes and the geochemical behaviour of that element. For example, Ni and Cr are compatible in olivine, and clinopyroxene and spinel-group minerals, respectively, and therefore these elements correlate well with MgO, TiO_2 and Fe (in this data set expressed as Fe_2O_3) in mafic lavas (Fig. 8). In comparison, chalcophile elements, such as Cu and especially the Pd-group PGE (PPGE; Rh, Pt, Pd), are controlled almost exclusively by sulphides, whether as sulphide liquid or base metal sulphides (BMS) and PGM (e.g., Naldrett, 2004; Keays and Lightfoot, 2007; Lorand and Luguet, 2016 and

525 references therein). Iridium-group PGE (IPGE; Os, Ir and Ru) although chalcophile, are also
526 compatible in spinel (especially chromite) and olivine (Barnes and Picard, 1993; Pitcher et al., 2009).
527 With higher degrees of partial melting, BMS will eventually be exhausted in the mantle source, and
528 spinel and olivine will begin to melt, therefore increasing the abundance of IPGE broadly in
529 correlation with increasing MgO, Ni, and Cr in the silicate magma produced. During ascent through
530 the lithosphere, this magma will undergo magmatic differentiation via fractional crystallisation of
531 olivine, pyroxene, spinel-group minerals and other silicate and oxide mineral phases, thereby further
532 modifying the concentration of MgO, Ni, Cr and the IPGE in combination with other major and trace
533 element proxies for fractionation and contamination.

534 If PCs can loosely be viewed as proxies for processes controlling element variability (e.g., Steiner et
535 al., 2019), in this case as a result of magmatic differentiation, we can begin to interpret the
536 elemental correlations consistently displayed through dimensionality reduction as constituents to
537 overall magma geochemical variability. PC1 (43.99% of data set variability) is likely to represent the
538 fractionation of olivine from the parental magma, given the close correlation with olivine-compatible
539 MgO, Ni, Cr and the IPGE in Figure 4b. PC2 (20.06% of data set variability), which primarily influences
540 TiO_2, Fe_2O_3, Cu and Pd concentrations (Fig. 4b), could represent a more complicated combination of
541 silicate, oxide and sulphide fractionation controls within parental magmas. For example, as mafic
542 magmas crystallise silicate minerals, the chalcophile elements (e.g., Cu and Pd) will become
543 increasingly concentrated in the residual melt until an immiscible sulphide liquid is exsolved
544 (following sulphide-saturation). At this point, chalcophiles will partition into the sulphide liquid,
545 depleting the remaining silicate melt for these elements (Naldrett, 2011 and references therein;
546 Ripley and Li, 2013). Meanwhile, the concentration of Ti in the magma will be affected by the
547 fractional crystallisation of silicate minerals (incorporating minor amounts of Ti; e.g., in
548 clinopyroxene) and oxide minerals (such as the spinel-group) depleting the residual liquid
549 composition *vs* potential addition of Ti to the magma via crustal contamination.

In the context of the NAIP, the three MLA-based geochemical end-members established from Figures 7 to 9 may be framed in terms of the evolving plume environment across this region. Geographic categories from Hughes et al. (2015) can be seen to ascribe to different dominant MLA-based end-members in Figures 10 to 12 (elemental concentrations are summarised in Table 2). In Figure 10, West Greenland has the highest proportion of Group 3 lavas of all localities (48%); East Greenland (onshore and offshore) has dominant Group 1 affiliations (> 70%); BPIP also has a prominent Group 1 affiliation (51%); and Iceland has a more equal split of each cluster.

Group 3 is the end-member most easily characterised within the NAIP setting given its distinctive strong association with the dominant PC1 and corresponding enrichment in olivine-compatible elements MgO, Ni, Cr and the IPGE (Fig. 4b, Fig. 5 and Fig. 12a). With MgO concentrations mainly between 15 and 25% (Fig. 8a), we suggest that Group 3 could represent higher degree partial melts of a mantle source, sequestering compatible elements like Ni and IPGE into the magma, followed by subsequent fractionation and accumulation of olivine (Fig. 11c). A sensible petrological NAIP association with this end-member would be the picrites temporally related to early NAIP lavas in West Greenlandic plateaus (e.g., the uncontaminated members of the Vaîgat Formation; Lightfoot et al., 1997), that are interpreted to have formed from magmas produced by up to 25% partial melting (Larsen and Pedersen, 2000; Andersen et al., 2002) – Figures 10 and 12. Vaîgat tholeiites record low $^{87}Sr/^{86}Sr$, high εNd and high γOs values consistent with asthenospheric plume melts with minimal SCLM interaction (Larsen et al., 2003).

Group 1 is primarily characterised by concentrations of MgO < 8.5% and TiO_2 ~2.0-2.5%, in addition to up to five times enrichment in Cu compared to other groups (Fig. 8c, Fig. 13 and Table 2), as expressed by PC2 (Fig. 4b). Figure 11c appears to suggest that the fractionation of elements incompatible in certain silicate minerals controls this group. The concentration of Cu will increase in fractionated magmas in the absence of a sulphide phase (Holwell et al., 2012 and references therein), and combined with earlier fractional crystallisation of olivine (or a similar mineral in which

575 Ti was incompatible), TiO_2 would be elevated in the Group 1 component. In plume environments,
576 higher TiO_2 concentrations are also thought to indicate magma erupting through thicker crust and
577 SCLM, which correlates with longer residence time, more differentiated/evolved compositions and
578 increased incorporation of high field strength and lithophile elements from crustal sources (e.g.,
579 Arndt et al., 1993; Gibson et al., 1995). We envisage the high-Ti basalts of East Greenland (such as
580 those in Kangertittivaq) could represent the product of this end-member as a more contaminated or
581 differentiated melt (Tegner et al., 1998), tying with Group 1 prominence in both East Greenland
582 categories from Hughes et al. (2015). Lavas from East Greenland have been noted to have high
583 $^{87}Sr/^{86}Sr$ and low εNd, often cited as an indication of lithospheric contamination (Kent and Fitton,
584 2000). However, it must be considered that due to consistency of data with PGE geochemistry
585 available to this study, we deal with eleven elements commonly used to document Ni-Cu-PGE
586 mineralisation and mafic-ultramafic magmatic systems; the data set does not include alkali,
587 lithophile or rare earth element concentrations. While TiO_2 is often a good initial indicator of mafic
588 magma contamination, interrogating a wider palette of lithophile elements could enhance clustering
589 and interpretations. Further testing on a larger element suite is necessary to confirm the link
590 between Group 1, TiO_2 and crustal contamination in combination with silicate-incompatible element
591 fractionation.

592 Group 2, the smallest group by sample number, does not exhibit the enrichments captured by
593 Groups 1 and 3, and almost always has the lowest range of concentrations for all variables (Fig. 13
594 and Table 2). As the intermediate end-member, it can be viewed in one of two ways – that it
595 represents a geochemically depleted source (although this is difficult to test without a wider
596 elemental suite) or, more likely, that it is a mixed source with no single dominant geochemical
597 control. If Groups 1 and 3 represent the end-members for higher amounts of incompatible
598 fractionation and higher degrees of olivine accumulation and/or partial melting in the NAIP
599 asthenosphere, respectively, Group 2 could simply represent the absence of strong multi-element
600 geochemical contributions from these factors. It could also represent a different higher-Fe_2O_3 (i.e.

previously melted) mantle source component (as per Korenaga & Kelemen, 2000). This end-member may relate to the North Atlantic End-Member (NAEM) proposed by Ellam & Stuart (2000), a ubiquitous 'background' component in NAIP melts defined by its distinctive Pb-isotopic signature through all NAIP lavas. It is possible that this end-member is derived mainly from depleted upper mantle material, which couples well with our multi-element assessment of the group.

An interesting feature presented by the MLA workflow is that categories of a similar age in the NAIP (and by extension belonging to a similar geodynamic setting) do not necessarily share the same high-dimensional data interpretation. The BPIP category shares more in common in terms of multi-element variability with later lavas in East Greenland or Iceland than it does with West Greenland despite being the same age (Fig. 1b). We would expect that BPIP lavas would share (i) a similar degree of partial melting and (ii) similar potential for crustal contamination to Greenlandic lavas, given the comparable geodynamics at this point in NAIP development (pre-rift continental flood basalts; Fig. 1b). Isotope data for both localities confirm their inherent link to the plume source (Saunders et al., 1997), so the discrepancy in cluster distribution must be associated with differences in the high-dimensional data structure as per their distinct positioning within t-SNE plots in Figure 12a-b. We suggest that, given BPIP is not defined by the absence of Groups 1 or 3 but simply reduced proportions of them, there are asymmetrical localised variations in magma sources on either side of the sampled hotspot, similar to other localities like the Tristan plume in the southern Atlantic (e.g., Peate, 1997; Hoernle et al., 2015; Rämö et al., 2016). In summary, the MLA end-members demonstrate that there are a combination of magmatic differentiation processes taking place in the NAIP system, and that the net effect of those processes has been to produce an array of lava compositions that fall largely into three categories in multi-dimensional (multi-element) space.

4.3 Platinum and palladium: Impacts on metal basket

625 The most prominent and significant finding from our study, the unique behaviour of specific PGE
626 within the NAIP high-dimensional data structure, demonstrates the utility of the MLA workflow for
627 exploring magmatic provinces from a 'mineral systems' perspective (e.g., McCuaig et al., 2010).
628 Using a holistic approach to metallogenic systems, we can examine regional fertility on different
629 scales and begin to understand the metal basket available to intraplate magmas.

630 It should be noted that Rh appears to fulfil a chemically intermediate role between IPGE and PPGE,
631 appearing between the two populations in Figure 4b and 4c; Rh also shows more distinct clustering
632 in Figure 8f. The mineralogical and chemical division of the IPGE and Rh from Pt and Pd is potentially
633 reflected in their separate contribution to NAIP data set variability - Figure 4b-d consistently
634 separates the two groups in terms of dominant PC influence. Controls on Pt and Pd concentrations
635 (particularly Pt) in NAIP lavas cannot be described as succinctly as the other nine elements, i.e. via
636 Groups 1 to 3 and their collective geochemical significance, and it is likely that instead of a major
637 dominant control linked to magmatic differentiation (e.g., PC1 or PC2) Pt and Pd are controlled by
638 smaller-scale processes that affect them exclusively. The workflow is able to effectively work with
639 major data trends while simultaneously identifying more subtle features extrinsic to these dominant
640 multi-element interpretations, which we can interrogate further in the framework of low-level or
641 localised geochemical controls within our system.

642 While PC1-2 have a combined contribution to Pt and Pd variability between the Group 1 and 3 end-
643 members, PC3-4 (which account for ~20% of data set variability; Fig. 4a) have a unique influence on
644 these two elements, far from the neat vector trends in the featured bivariate plots. Hence the
645 controls on Pt and Pd concentrations appear to be more complex and cryptic than other modelled
646 elements in PCA (Fig. 4b-d). This is further evidenced by the digression from significant patterns in
647 the subsequent stages of the workflow - firstly in the embedding created via t-SNE, which shows Pt
648 and Pd variability contradicting the major bimodal trends established by the other elements (Fig. 5);
649 and secondly in *k*-means clustering in all attempted model set-ups, which fails to capture distinct

650 populations in these two variables (Fig. 9a). Consistently high Pt/Pd signatures are found in BPIP and
651 West Greenland lavas (Hughes et al., 2015) yet these two localities exhibit different multi-element
652 variability trends in the rest of the MLA results, clearly indicating that the geographic Pt/Pd trend is
653 more nuanced than other grouped elements captured within the data structure (as in Fig. 5).

654 Platinum is visibly separated in Figure 4c from all other elements (even Pd and Rh) implying that
655 whatever control PC3-4 represents, Pt is the element chiefly affected. In comparison, Pd, while
656 lacking affinity for multi-element classification models, does correlate moderately with Cu (Fig. 3)
657 likely on account of similarities in the mineralogical controls on their fractionation. Together with t-
658 SNE embedding (Fig. 5) and clusters outside the dominant multi-element categories (Fig. 8 and 9),
659 the independent variation of Pt appears to be a small but significant control.

660 As illustrated, multi-element geochemistry can vary locally as a function of the components available
661 to the melting environment (Fig. 10). For example, the Hebridean Basin in western Scotland could
662 provide large amounts of sedimentary material via contamination during magma ascent, different
663 from those available to magmas ascending through the Greenlandic lithosphere. These mid- to
664 shallow-crustal contaminants provide effective localised signatures, changing the resulting
665 geochemistry of melts and potentially their metal basket (Andersen et al., 2002), which is of
666 particular importance to our work in the form of PGE prospectivity. In the absence of a full set of
667 major, trace and precious element data for all localities, we can use metallogenic studies of the
668 region to inform our assessment of Pt and Pd behaviour, specifically addressing the independent
669 roles of the elements in high-dimensional space as resolved by MLA and the apparent localised
670 controls on their enrichment. The 'nugget effect' of discontinuous geological occurrence (e.g.,
671 Dominy et al., 2003) is common for metals with complex geochemical and physical partitioning
672 behaviour, such as Au, base metals and the PGE, emphasising the importance of recognising spatial
673 and temporal variations in magma enrichment detached from major element oxides like MgO, Fe_2O_3
674 and TiO_2. Gold has been seen to vary locally in lavas from Iceland and its adjoining Reykjanes Ridge,

675 in relation to sulphide saturation conditions and differing mantle sources of the plume-derived
676 melts, rather than as a direct function of the degree of partial melting (Webber et al., 2013). Webber
677 et al. (2013) therefore illustrate the impact of source heterogeneity in regional metal basket
678 variation and metallogenesis in a manner analogous to magma isotopic signatures transgressing
679 petrological classifications. Additionally, Au, Co, Cu and the PGE amongst other elements have been
680 speculated to be locally enriched via metasomatism into the lithospheric mantle (e.g., Mitchell and
681 Keays, 1981; Tassara et al., 2017). Metal enrichment of the SCLM through subduction-related
682 volatile and fluid transportation is widely documented, and Pt, Pd and Au are particularly thought to
683 be mobilised in these metasomatic environments (e.g., Hughes et al., 2017; Tassara et al., 2017;
684 Holwell et al., 2019; Choi et al., in press). We envisage that metasomatism (or similar hydro-
685 magmatic processes) could have resulted in the 'pre-conditioning' of the SCLM above the proto-
686 Icelandic plume, leading to its heterogeneous enrichment in Pt and a shift in local metal basket. Such
687 controls evidently do not map onto our 3 MLA-based multi-element end-members, reinforcing the
688 potential localised nature of high concentrations within the broader system.

689 The complexities of localised PGE behaviour and partitioning during partial melting of the mantle has
690 been studied experimentally and empirically (e.g., Keays, 1982; Hamlyn and Keays, 1986; Peach et
691 al., 1990; Rehkämper et al., 1997; Lorand et al., 1999; Ballhaus et al., 2001; Lorand and Alard, 2001;
692 Luguet et al., 2003; Bockrath et al., 2004; Righter et al., 2004; Pitcher et al., 2009; Locmelis et al.,
693 2013; Lorand et al., 2013; Mungall and Brenan, 2014; Lorand and Luguet, 2016; Luguet and Reisberg,
694 2016). IPGE are considered compatible during partial melting, with an affiliation for some silicate
695 and oxide minerals, and thus behave in a similar fashion to Ni and Cr (e.g., Brenan and Andrews,
696 2001; Maier et al., 2003; Pitcher et al., 2009; Pagé et al., 2012). On the other hand, PPGE (especially
697 Pd) are largely hosted by BMS (e.g., Mitchell and Keays, 1981; Alard et al., 2000; Lorand and Luguet,
698 2016; Luguet and Reisberg, 2016 and references therein) and are incompatible in silicates and oxides
699 (e.g., Hill et al., 2000; Righter et al., 2004). Melting-induced fractionation of PGE may also stem from
700 the coexistence of two mantle sulphide phases in the mantle source: a crystalline monosulphide

enriched in Fe, Ni and IPGE *vs* an immiscible sulphide enriched in Cu, Ni and PPGE (e.g., Ballhaus et al., 2001; Lorand et al., 2013; Lorand and Luguet, 2016; Luguet and Reisberg, 2016). Significant fractionation between PPGE and IPGE can be caused by low degree partial melting as the immiscible sulphide becomes mobilised into silicate magma, leaving a monosulphide residue. With increasing partial melting, sulphide phases, including the monosulphide, may be exhausted and silicates and oxides (such as olivine and chromite) will also begin to release their IPGE budget (e.g., Keays, 1982; Rehkämper et al., 1997; Maier et al., 2003; Mungall and Brenan, 2014). Accordingly, komatiites have much lower Pd/Ir ratios than basalts (e.g., Rehkämper et al., 1999).

Fractionation between individual elements of the PPGE may theoretically be possible during partial melting on account of their extremely high sulphide-silicate partitioning coefficients (D_{Pt} = 317,000 *vs* D_{Pd} = 190,000 e.g., Mungall & Brenan, 2014). However, such strong chalcophile behaviour of both elements (extremely high partition coefficients, D) mean that this fractionation effect is likely to be marginal at best, and dwarfed in comparison to the efficiency of sulphide melting and extraction from the mantle – even a small residue of sulphide in the source can significantly inhibit the concentration of PGE in the silicate melt generated. It has been suggested that repeated partial melting events of the same source region record increasing Pt/Pd ratios via early depletion in Pd (e.g., Keays, 1982; Keays et al., 1982; Hamlyn et al., 1985). Further, of the PPGE, Pt in particular can also occur in PGM and these may behave differently from BMS during partial melting (depending on whether these are as PGM-sulphides or alloys). Regardless of the precise mechanisms for Pt and Pd decoupling, crucial empirical evidence from mantle peridotite xenoliths indicates that there may be regional and local variations in the composition of mantle BMS and PGM and thus domains particularly enriched in Pt (Wittig et al., 2010; Hughes et al., 2017). For example, Hughes et al. (2017) found that mantle xenoliths from the keel of the Scottish portion of the North Atlantic Craton (which underlies portions of the BPIP) contain BMS enriched in PGE and characteristically bear micron-scale Pt-sulphides (cooperite). In context with the shift in NAIP lava Pt/Pd ratios through time, Hughes et al. (2015) suggested that the high Pt/Pd of earlier NAIP lavas resulted from entrainment of these Pt-

727 sulphides, thereby 'spiking' the composition of the asthenosphere-derived mantle plume magmas
728 during their ascent to the crust.

729 Ultimately, the MLA approach to geochemical data analysis adopted in this study provides robust
730 evidence that Pt (and to some extent Pd) is truly decoupled from the rest of the major and trace
731 element geochemistry of the NAIP LIP. The multi-elemental assessment of the data set is essentially
732 an integrated account of multiple geochemical processes (via magmatic differentiation; e.g., Fig. 2a)
733 affecting the final concentrations of elements in NAIP lavas, the cumulative result of which produces
734 Groups 1 to 3. The Pt/Pd shift in the NAIP lavas with time is more complex than a systematic
735 geographic variation; Pt is evidently being added to partial melts from a reservoir outwith the main
736 process of magma generation via asthenospheric partial melting, and we suggest the potential role
737 of a locally metasomatised SCLM in this process. The use of MLA demonstrates a mineral systems
738 approach to mantle source fertility and the manner in which the methodology can isolate specific
739 nonconforming trends in high-dimensional space.

740

741 **5 Conclusions**

742 The benefits of using machine learning to explore high dimensional data sets have been clearly
743 outlined by our study of NAIP lava geochemistry. By using a novel combined data science and
744 classical approach to bulk geochemical data sets, we have identified some key contributions to the
745 investigation into PGE metallogenic controls in plume environments:

746 1. PCA, t-SNE and *k*-means clustering identified major multi-element trends in NAIP lava
747 geochemistry and established three distinct geochemical end-members in the data set,
748 which represent the net result of magmatic differentiation processes in the geodynamic
749 system.

2. The workflow captured consistent variability information in nine of the eleven included variables, importantly isolating Pt and Pd as exceptional to end-member trends in all methods and input formations, implying a localised geochemical control for these two metals in the NAIP.
3. We suggest a locally metasomatised and Pt-sulphide-rich SCLM reservoir being incorporated into plume melts as a reasonable explanation for the unique Pt/Pd variability, based on previous findings regarding asthenospheric heterogeneity in the NAIP.
4. The role of feature extraction via dimensionality reduction is important for manageable high-dimensional geochemical investigations as illustrated by consistently improved clustering performance using Principal Components as input variables for MLAs.
5. Further advances in understanding this wide topic may be possible by integrating more elements in larger data sets from other plume regions (where PGE are of significant interest), particularly lithophile elements.

765 **Acknowledgements**

766 The authors would like to thank Chris Hawkesworth for sharing his insights, thorough discussion, and
767 for being supportive of this new approach to geochemistry at each step of the process. We would
768 also like to thank Matthew Head for his assistance with initial coding and Benedikt Steiner for his
769 help in PCA interpretation. We thank the two anonymous reviewers of the original submission of this
770 manuscript for their comments and feedback, and Kristoffer Szilas for his editorial handling. Finally,
771 we would like to thank University of Exeter's Vice Chancellor Scholarship for funding JJL's PhD.

772

773 **References**

774 Alard O., Griffin W. L., Lorand J. P., Jackson S. E. and O'Reilly S. Y. (2000) Non-chondritic distribution
775 of the highly siderophile elements in mantle sulphides. *Nature* **407**, 891–894.

776 Andersen J. C. Ø., Power M. R. and Momme P. (2002) Platinum-Group Elements in the Palaeogene
777 North Atlantic Igneous Province. The geology, geochemistry, mineralogy, and mineral
778 beneficiation of platinum-group elements. *L. J. Cabri. Montréal, Québec, Can. Inst. Mining,*
779 *Metall. Pet.* **CIM Spec.**, 637–667.

780 Andersen J. C. Ø., Rasmussen H., Nielsen T. F. D. and Rønsbo J. G. (1998) The Triple Group and the
781 Platinova gold and palladium reefs in the Skaergaard Intrusion: stratigraphic and petrographic
782 relations. *Econ. Geol.* **93**, 488–509.

783 Arndt N. T., Czamanske G. K., Wooden J. L. and Fedorenko V. A. (1993) Mantle and crustal
784 contributions to continental flood volcanism. *Tectonophysics* **223**, 39–52.

785 Balamurali M. and Melkumyan A. (2016) t-SNE Based Visualisation and Clustering of Geological
786 Domain. In *International Conference on Neural Information Processing*

787 Ballhaus C., Tredoux M. and Späth A. (2001) Phase relations in the Fe-Ni-Cu-PGE-S system at
788 magmatic temperature and application to massive sulphide ores of the the sudbury igneous
789 complex. *J. Petrol.* **42**, 1911–1926.

790 Barnes S. J. and Picard C. P. (1993) The behaviour of platinum-group elements during partial melting,
791 crystal fractionation, and sulphide segregation: An example from the Cape Smith Fold Belt,
792 northern Quebec. *Geochim. Cosmochim. Acta* **57**, 79–87.

793 Bercovici D. and Kelly A. (1997) The non-linear initiation of diapirs and plume heads. *Phys. Earth*
794 *Planet. Inter.* **101**, 119–130.

795 Berggren W. A., Kent D. V., Swisher C. C. and Aubry M. P. (1995) A revised Cenozoic geochronology
796 and chronostratigraphy. *Geochronol. Time Scales Glob. Stratigr. Correl. SEPM Spec. Publ.* **54**,
797 129–212.

798 Bockrath C., Ballhaus C. and Holzheid A. (2004) Fractionation of the platinum-group elements during
799 mantle melting. *Science.* **305**, 1951–1953.

800 Brenan J. M. and Andrews D. (2001) High-temperature stability of laurite and Ru-Os-Ir alloy and their
801 role in PGE fractionation in mafic magmas: Erratum. *Can. Mineral.* **39**, 341–360.

802 Butcher A. R., Pirrie D., Prichard H. M. and Fisher P. C. (1999) Platinum-group mineralization in the
803 Rum layered intrusion, Scottish Hebrides, UK. *J. Geol. Soc. London.* **156**, 213–216. Available at:
804 http://jgs.lyellcollection.org/cgi/doi/10.1144/gsjgs.156.2.0213.

805 Campbell I. H. and Griffiths R. W. (1990) Implications of mantle plume structure for the evolution of
806 flood basalts. *Earth Planet. Sci. Lett.* **99**, 79–93.

807 Chang W. C. (1983) On Using Principal Components Before Separating a Mixture of Two Multivariate
808 Normal Distributions. *Appl. Stat.* **32**, 267–275.

809 Choi, E., Fiorentini, M.L., Hughes, H.S.R., Guilani, A. (accepted). Platinum-group element and Au
810 geochemistry of Late Archean to Proterozoic calc-alkaline and alkaline magmas in the Yilgarn
811 Craton, Western Australia. *Lithos*. In press.

812 Cone K. A., Palin R. M. and Singha K. (2020) Unsupervised machine learning with petrological
813 database ApolloBasaltDB reveals complexity in lunar basalt major element oxide and mineral

distribution patterns. *Icarus*. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0019103520301731.

Courtillot V., Davaille A., Besse J. and Stock J. (2003) Three distinct types of hotspots in the Earth's mantle. *Earth Planet. Sci. Lett.* **205**, 295–308. Available at: /Users/cbeghein/Documents/PDFs/2003_Courtillot_Earth_and_Planetary_Science_Le.pdf%5Cn http://dx.doi.org/10.1016/S0012-821X(02)01048-8.

Davies D. L. and Bouldin D. W. (1979) A Cluster Separation Measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI**-**1**, 224–227.

Davis J. C. (2002) *Statistics and Data Analysis in Geology*. 3rd ed., John Wiley & Sons Inc.

Dominy S. C., Platten I. M. and Raine M. D. (2003) Grade and geological continuity in high-nugget effect gold-quartz reefs: Implications for resource estimation and reporting. *Trans. Inst. Min. Metall. Sect. B Appl. Earth Sci.* **112**, 239–259.

Ellam R. M. and Stuart F. M. (2000) The sub-lithospheric source of North Atlantic basalts: Evidence for, and significance of, a common end-member. *J. Petrol.* **41**, 919–932.

Farnham I. M., Singh A. K., Stetzenbach K. J. and Johannesson K. H. (2002) Treatment of nondetects in multivariate analysis of groundwater geochemistry data. *Chemom. Intell. Lab. Syst.* **60**, 265–281.

Fitton J., Saunders A., Norry M., Hardarson B. and Taylor R. (1997) Thermal and chemical structure of the Iceland plume. *Earth Planet. Sci. Lett. Planet. Sci. Lett.* **153**, 197–208. Available at: http://www.sciencedirect.com/science/article/pii/S0012821X97001702%5Cnhttp://ac.els-cdn.com/S0012821X97001702/1-s2.0-S0012821X97001702-main.pdf?_tid=67da9ecc-2ffc-11e3-bf1f-00000aab0f6c&acdnat=1381224867_3a9432c73a0c20384a002386889c18d7.

Fleet M. E., Crocket J. H., Liu M. and Stone W. E. (1999) Laboratory partitioning of platinum-group elements (PGE) and gold with application to magmatic sulfide-PGE deposits. *Lithos* **47**, 127–142.

Fleet M. E., Crocket J. H. and Stone W. E. (1996) Partitioning of platinum-group elements (Os, Ir, Ru, Pt, Pd) and gold between sulfide liquid and basalt melt. *Geochim. Cosmochim. Acta* **60**, 2397–2412.

Foulger G. R., Natland J. H. and Anderson D. L. (2005) A source for Icelandic magmas in remelted Iapetus crust. *J. Volcanol. Geotherm. Res.* **141**, 23–44. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0377027304003312.

Ghannadpour S., Hezarkhani A. and Farahbakhsh E. (2013) An Investigation of Pb Geochemical Behavior Respect to Those of Fe and Zn Based on k- Means Clustering Method. *J. Tethys* **1**, 291–302.

Gibson S. A., Thompson R. N., Dickin A. P. and Leonardos O. H. (1995) High-Ti and low-Ti mafic potassic magmas: Key to plume-lithosphere interactions and continental flood-basalt genesis. *Earth Planet. Sci. Lett.* **136**, 149–165.

Griffiths R. W. and Campbell I. H. (1990) Stirring and structure in mantle starting plumes. *Earth Planet. Sci. Lett.* **99**, 66–78.

Hamlyn P. R. and Keays R. R. (1986) Sulfur saturation and second-stage melts: application to the Bushveld platinum metal deposits. *Econ. Geol.* **81**, 1431–1445.

Hamlyn P. R., Keays R. R., Cameron W. E., Crawford A. J. and Waldron H. M. (1985) Precious metals in magnesian low-Ti lavas: Implications for metallogenesis and sulfur saturation in primary

magmas. *Geochim. Cosmochim. Acta* **49**, 1797–1811.

Hastie T., Tibshirani R. and Friedman J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed., Springer.

Hawkesworth C. and Scherstén A. (2007) Mantle plumes and geochemistry. *Chem. Geol.* **241**, 319–331.

Hill E., Wood B. J. and Blundy J. D. (2000) The effect of Ca-Tschermaks component on trace element partitioning between clinopyroxene and silicate melt. *Lithos* **53**, 203–215.

Hoernle K., Rohde J., Hauff F., Garbe-Schönberg D., Homrighausen S., Werner R. and Morgan J. P. (2015) How and when plume zonation appeared during the 132 Myr evolution of the Tristan Hotspot. *Nat. Commun.* **6**, 7799. Available at: http://www.nature.com/articles/ncomms8799.

Hole M. J. and Natland J. H. (2019) Magmatism in the North Atlantic Igneous Province; mantle temperatures, rifting and geodynamics. *Earth-Science Rev.*, 1–24. Available at: https://doi.org/10.1016/j.earscirev.2019.02.011.

Holwell D. A., Abraham-James T., Keays R. R. and Boyce A. J. (2012) The nature and genesis of marginal Cu-PGE-Au sulphide mineralisation in Paleogene Macrodykes of the Kangerlussuaq region, East Greenland. *Miner. Depos.* **47**, 3–21.

Holwell D. A., Fiorentini M., McDonald I., Lu Y., Giuliani A., Smith D. J., Keith M. and Locmelis M. (2019) A metasomatized lithospheric mantle control on the metallogenic signature of post-subduction magmatism. *Nat. Commun.* **10**, 1–10. Available at: http://dx.doi.org/10.1038/s41467-019-11065-4.

Horni J. Á. ., Hopper J. R., Blischke A., Geisler W. H., Stewart M., McDermott K., Judge M., Erlendsson Ö. and Árting U. (2017) Regional distribution of volcanism within the North Atlantic Igneous Province. *Geol. Soc. London, Spec. Publ.* **447**, 105–125. Available at: http://sp.lyellcollection.org/lookup/doi/10.1144/SP447.18.

Horrocks T., Holden E. J., Wedge D., Wijns C. and Fiorentini M. (2019) Geochemical characterisation of rock hydration processes using t-SNE. *Comput. Geosci.* **124**, 46–57. Available at: https://doi.org/10.1016/j.cageo.2018.12.005.

Hotelling H. (1933) Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **25**, 417–441.

Howarth G. H. and Harris C. (2017) Discriminating between pyroxenite and peridotite sources for continental flood basalts (CFB) in southern Africa using olivine chemistry. *Earth Planet. Sci. Lett.* **475**, 143–151. Available at: http://dx.doi.org/10.1016/j.epsl.2017.07.043.

Howarth R. J. (1983) Statistics and Data Analysis in Geochemical Prospecting. In *Handbook ofExploration Geochemistry* (ed. G. J. S. Govett). Elsevier, Amsterdam.

Hughes H. S. R., McDonald I., Goodenough K. M., Ciborowski T. J. R., Kerr A. C., Davies J. H. F. L. and Selby D. (2014) Enriched lithospheric mantle keel below the Scottish margin of the North Atlantic Craton: Evidence from the Palaeoproterozoic Scourie Dyke Swarm and mantle xenoliths. *Precambrian Res.* **250**, 97–126. Available at: http://dx.doi.org/10.1016/j.precamres.2014.05.026.

Hughes H. S. R., McDonald I. and Kerr A. C. (2015) Platinum-group element signatures in the North Atlantic Igneous Province: Implications for mantle controls on metal budgets during continental breakup. *Lithos* **233**, 89–110. Available at: http://dx.doi.org/10.1016/j.lithos.2015.05.005.

Hughes H. S. R., McDonald I., Loocke M., Butler I. B., Upton B. G. J. and Faithfull J. W. (2017)

Paradoxical co-existing base metal sulphides in the mantle: The multi-event record preserved in Loch Roag peridotite xenoliths, North Atlantic Craton. *Lithos* **276**, 103–121. Available at: http://dx.doi.org/10.1016/j.lithos.2016.09.035.

Hyvärinen A., Karhunen J. and Oja E. (2001) Introduction. In *Independent Component Analysis* (ed. S. Haykin). John Wiley & Sons, Inc., New York.

Iwamori H., Yoshida K., Najamura H., Kuwatani T., Hamada M., Haraguchi S. and Ueki K. (2017) Classification of geochemical data based on multivariate statistical analyses: Complementary roles of cluster, principal component, and independent component analyses. *Geochemistry, Geophys. Geosystems* **18**, 994–1012.

Jellinek A. M. and Manga M. (2004) Links Between Long-lived Hot Spots, Mantle Plumes, D", And Plate Tectonics. *Rev. Geophys.* **42**, 1–35. Available at: http://www.seismo.berkeley.edu/~manga/paper72.pdf.

Jolliffe I. T. (2002) *Principal Component Analysis*. 2nd ed., Springer, New York.

Keays R. R. (1982) Palladium and iridium in komatiites and associated rocks: application to petrogenetic problems. In *Komatiites* (eds. N. T. Arndt and E. G. Nisbet). George Allen and Unwin, London. pp. 435–455.

Keays R. R. and Lightfoot P. C. (2007) Siderophile and chalcophile metal variations in Tertiary picrites and basalts from West Greenland with implications for the sulphide saturation history of continental flood basalt magmas. *Miner. Depos.* **42**, 319–336.

Keays R. R., Nickel E. H., Groves D. I. and McGoldrick P. J. (1982) Iridium and Palladium as Discriminants of Volcanic-Exhalative, Hydrothermal, and Magmatic Nickel Sulfide Mineralization. *Econ. Geol.* **77**, 1535–1547.

Kellogg L. H. and King S. D. (1993) Effect of mantle plumes on the growth of D” by reaction between the core and mantle. *Geophys. Res. Lett.* **20**, 379–382.

Kent R. a Y. W. and Fitton J. G. (2000) Mantle Sources and Melting Dynamics in the British Palaeogene Igneous Province. *J. Petrol.* **41**, 1023–1040.

Korenaga J. and Kelemen P. B. (2000) Major element heterogeneity in the mantle source of the North Atlantic igneous province. *Earth Planet. Sci. Lett.* **184**, 251–268.

Kreyszig E. (1979) *Advanced Engineering Mathematics*. 4th ed., Wiley.

Krijthe J. H. (2015) Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation. Available at: https://github.com/jkrijthe/Rtsne [Accessed April 30, 2020].

Kullback S. and Leibler R. A. (1959) *Information Theory and Statistics*. 3rd ed., John Wiley & Sons Inc. Available at: http://arxiv.org/abs/1706.01538.

Larsen L. M. and Pedersen A. K. (2000) Processes in high-Mg, high-T magmas: Evidence from olivine, chromite and glass in palaeogene picrites from West Greenland. *J. Petrol.* **41**, 1071–1098.

Larsen L. M., Pedersen A. K., Sundvoll B. and Frei R. (2003) Alkali picrites formed by melting of old metasomatized lithospheric mantle: Manîtdlat member, vaigat formation, Palaeocene of West Greenland. *J. Petrol.* **44**, 3–38.

Lawver L. A. and Muller R. D. (1994) Iceland hotspot track. *Geology* **22**, 311–314.

Lightfoot P. C., Hawkesworth C. J., Olshefsky K., Green T., Doherty W. and Keays R. R. (1997) Geochemistry of Tertiary tholeiites and picrites from Qeqertarssuaq (Disko Island) and

941 Nuussuaq, West Greenland with implications for the mineral potential of comagmatic
942 intrusions. *Contrib. to Mineral. Petrol.* **128**, 139–163.

943 Locmelis M., Fiorentini M. L., Barnes S. J. and Pearson N. J. (2013) Ruthenium Variation in Chromite
944 from Komatiites and Komatiitic Basalts—A Potential Mineralogical Indicator for Nickel Sulfide
945 Mineralization. *Econ. Geol.* **108**, 355–364.

946 Lorand J. P. and Alard O. (2001) Platinum-group element abundances in the upper mantle: New
947 constraints from in situ and whole-rock analyses of massif central xenoliths (France). *Geochim.*
948 *Cosmochim. Acta* **65**, 2789–2806.

949 Lorand J. P. and Luguet A. (2016) Chalcophile and siderophile elements in mantle rocks: Trace
950 elements controlled by trace minerals. *Rev. Mineral. Geochemistry* **81**, 441–488. Available at:
951 http://rimg.geoscienceworld.org/lookup/doi/10.2138/rmg.2016.81.08.

952 Lorand J. P., Luguet A. and Alard O. (2013) Platinum-group element systematics and petrogenetic
953 processing of the continental upper mantle: A review. *Lithos* **164–167**, 2–21. Available at:
954 http://dx.doi.org/10.1016/j.lithos.2012.08.017.

955 Lorand J. P., Pattou L. and Gros M. (1999) Fractionation of Platinum-group elements and gold in the
956 upper mantle: A detailed study in Pyrenean orogenic lherzolites. *J. Petrol.* **40**, 957–981.

957 Luguet A., Lorand J. P. and Seyler M. (2003) Sulfide petrology and highly siderophile element
958 geochemistry of abyssal peridotites: A coupled study of samples from the Kane Fracture Zone
959 (45°W 23°20N, MARK area, Atlantic Ocean). *Geochim. Cosmochim. Acta* **67**, 1553–1570.

960 Luguet Ambre and Reisberg L. (2016) Highly Siderophile Element and 187Os Signatures in Non-
961 cratonic Basalt-hosted Peridotite Xenoliths: Unravelling the Origin and Evolution of the Post-
962 Archean Lithospheric Mantle. *Rev. Mineral. Geochemistry* **81**, 305–367. Available at:
963 http://rimg.geoscienceworld.org/content/81/1/305.short.

964 Luguet A. and Reisberg L. (2016) Highly Siderophile Element and 187Os Signatures in Non-cratonic
965 Basalt-hosted Peridotite Xenoliths: Unravelling the Origin and Evolution of the Post-Archean
966 Lithospheric Mantle. *Rev. Mineral. Geochemistry* **81**, 305–367.

967 van der Maaten L. (2014) Accelerating t-SNE using tree-based algorithms. *J. Mach. Learn. Res.* **15**,
968 3221–3245.

969 van der Maaten L. and Hinton G. (2008) Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–
970 2605.

971 MacQueen J. (1967) Some methods for classification and analysis of multivariate observations. *Proc.*
972 *Fifth Berkeley Symp. Math. Stat. Probab.* **1**, 281–297.

973 Maier W. D. and Groves D. I. (2011) Temporal and spatial controls on the formation of magmatic PGE
974 and Ni-Cu deposits. *Miner. Depos.* **46**, 841–857.

975 Maier W. D., Roelofse F. and Barnes S.-J. (2003) The Concentration of the Platinum-Group Elements
976 in South African Komatiites: Implications for Mantle Sources, Melting Regime and PGE
977 Fractionation during Crystallization. *J. Petrol.* **44**, 1787–1804.

978 Marsland S. (2009) *Machine Learning An Algorithmic Perspective Second Edition*. 2nd ed. eds. R.
979 Herbrich and T. Graepel, CRC Press.

980 Martín-Fernández J. A., Hron K., Templ M., Filzmoser P. and Palarea-Albaladejo J. (2012) Model-
981 based replacement of rounded zeros in compositional data: Classical and robust approaches.
982 *Comput. Stat. Data Anal.* **56**, 2688–2704. Available at:
983 https://linkinghub.elsevier.com/retrieve/pii/S0167947312000941.

McCuaig T. C., Beresford S. and Hronsky J. (2010) Translating the mineral systems approach into an effective exploration targeting system. *Ore Geol. Rev.* **38**, 128–138. Available at: http://dx.doi.org/10.1016/j.oregeorev.2010.05.008.

McDonough W. F. and Sun S. S. (1995) The composition of Earth. *Chem. Geol.* **120**, 223–253.

McKenzie D. and White R. (1989) Magmatism at rift zones: The generation of volcanic continental margins and flood basalts. *J. Geophys. Res.* **94**, 7685–7729.

Michie D., Spiegelhalter D. J. and Taylor C. C. (1994) *Machine Learning, Neural and Statistical Classification*., Ellise Horwood Limited.

Mitchell R. H. and Keays R. R. (1981) Abundance and distribution of gold, palladium and iridium in some spinel and garnet lherzolites: implications for the nature and origin of precious metal-rich intergranular components in the upper mantle. *Geochim. Cosmochim. Acta* **45**, 2425–2442.

Momme P., Óskarsson N. and Keays R. R. (2003) Platinum-group elements in the Icelandic rift system: melting processes and mantle sources beneath Iceland. *Chem. Geol.* **196**, 209–234. Available at: https://linkinghub.elsevier.com/retrieve/pii/S000925410200414X.

Momme P., Tegner C., Brooks C. K. and Keays R. R. (2002) The behaviour of platinum-group elements in basalts from the East Greenland rifted margin. *Contrib. to Mineral. Petrol.* **143**, 133–153.

Mungall J. E. and Brenan J. (2014) Partitioning of platinum-group elements and Au between sulfide liquid and basalt and the origins of mantle-crust fractionation of the chalcophile elements. *Geochim. Cosmochim. Acta* **125**, 265–289. Available at: http://dx.doi.org/10.1016/j.gca.2013.10.002.

Naldrett A. J. (2011) Fundamentals of Magmatic Sulfide Deposits. *Rev. Econ. Geol.* **17**, 1–50.

Naldrett A. J. (2004) *Magmatic Sulfide Deposits*., Available at: http://www.minersoc.org/pages/Archive-MM/Volume_54/54-377-675.pdf.

Nguyen L. H. and Holmes S. (2019) Ten quick tips for effective dimensionality reduction. *PLoS Comput. Biol.* **15**, 1–19.

Pagé P., Barnes S. J., Bédard J. H. and Zientek M. L. (2012) In situ determination of Os, Ir, and Ru in chromites formed from komatiite, tholeiite and boninite magmas: Implications for chromite control of Os, Ir and Ru during partial melting and crystal fractionation. *Chem. Geol.* **302–303**, 3–15. Available at: http://dx.doi.org/10.1016/j.chemgeo.2011.06.006.

Peach C. L., Mathez E. A. and Keays R. R. (1990) Sulfide melt-silicate melt distribution coefficients for noble metals and other chalcophile elements as deduced from MORB: Implications for partial melting. *Geochim. Cosmochim. Acta* **54**, 3379–3389.

Pearson K. (1901) On lines and planes of closest fit to systems of points in space. *Phil. Mag. J. Sci.* **2**, 559–572.

Peate D. W. (1997) The Paraná-Etendeka province. *Geophys. Monogr. Ser. Ser* **100**, 217–245.

Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M. and Duchesnay E. (2011) Scikit-learn: Machine Learning in Python. *J. Ma* **12**, 2825–2830.

Philipp H., Eckhardt J.-D. and Puchelt H. (2001) Platinum-Group Elements (PGE) in Basalts of the Seaward-Dipping Reflector Sequence, SE Greenland Coast. *J. Petrol.* **42**, 407–432.

Pirajno F. and Santosh M. (2015) Mantle plumes, supercontinents, intracontinental rifting and

1025 mineral systems. *Precambrian Res.* **259**, 243–261. Available at:
1026 http://dx.doi.org/10.1016/j.precamres.2014.12.016.

1027 Pirrie D., Power M. R., Andersen J. C. Ø. and Butcher a. R. (2000) Platinum-group mineralization in
1028 the Tertiary Igneous Province: new data from Mull and Skye, Scottish Inner Hebrides, UK. *Geol.*
1029 *Mag.* **137**, 651–658.

1030 Pitcher L., Helz R. T., Walker R. J. and Piccoli P. (2009) Fractionation of the platinum-group elments
1031 and Re during crystallization of basalt in Kilauea Iki Lava Lake, Hawaii. *Chem. Geol.* **260**, 196–
1032 210. Available at: http://dx.doi.org/10.1016/j.chemgeo.2008.12.022.

1033 Rämö O. T., Heikkilä P. A. and Pulkkinen A. H. (2016) Geochemistry of Paraná-Etendeka basalts from
1034 Misiones, Argentina: Some new insights into the petrogenesis of high-Ti continental flood
1035 basalts. *J. South Am. Earth Sci.* **67**, 25–39.

1036 Rehkämper M., Halliday A. N., Barfod D., Fitton J. G. and Dawson J. B. (1997) Platinum-group
1037 element abundance patterns in different mantle environments. *Science* **278**, 1595–1598.

1038 Rehkämper M., Halliday A. N., Fitton J. G., Lee D. C., Wieneke M. and Arndt N. T. (1999) Ir, Ru, Pt,
1039 and Pd in basalts and komatiites: New constraints for the geochemical behavior of the
1040 platinum-group elements in the mantle. *Geochim. Cosmochim. Acta* **63**, 3915–3934.

1041 Righter K., Campbell A. J., Humayun M. and Hervig R. L. (2004) Partitioning of Ru, Rh, Pd, Re, Ir, and
1042 Au between Cr-bearing spinel, olivine, pyroxene and silicate melts. *Geochim. Cosmochim. Acta*
1043 **68**, 867–880.

1044 Ripley E. M. and Li C. (2013) Sulfide saturation in mafic magmas: Is external sulfur required for
1045 magmatic Ni-Cu-(PGE) ore genesis? *Econ. Geol.* **108**, 45–58.

1046 Saunders A. D., Fitton J. G., Kerr A. C., Norry M. J. and Kent R. W. (1997) The North Atlantic Igneous
1047 Province. In *Large Igneous Provinces: Continental, Oceanic and Planetary Flood Volcanism* (eds.
1048 J. J. Mahoney and M. . Coffin). American Geophysical Union Monograph. pp. 45–93. Available
1049 at: http://doi.wiley.com/10.1029/GM100p0045.

1050 Steiner B. M., Rollinson G. K. and Condron J. M. (2019) An exploration study of the Kagenfels and
1051 Natzwiller granites, Northern Vosges Mountains, France: a combined approach of stream
1052 sediment geochemistry and automated mineralogy. *Minerals* **9**, 1–28.

1053 Tassara S., González-Jiménez J. M., Reich M., Schilling M. E., Morata D., Begg G., Saunders E., Griffin
1054 W. L., O'Reilly S. Y., Grégoire M., Barra F. and Corgne A. (2017) Plume-subduction interaction
1055 forms large auriferous provinces. *Nat. Commun.* **8**, 1–7.

1056 Taylor G. J., Martel L. M. V., Karunatillake S., Gasnault O. and Boynton W. V. (2010) Mapping Mars
1057 geochemically. *Geology* **38**, 183–186.

1058 Tegner C., Duncan R. A., Bernstein S., Brooks C. K., Bird D. K. and Storey M. (1998) 40Ar-39Ar
1059 geochronology of Tertiary mafic intrusions along the East Greenland rifted margin: Relation to
1060 flood basalts and the iceland hotspot track. *Earth Planet. Sci. Lett.* **156**, 75–88.

1061 Trela J., Vidito C., Gazel E., Herzberg C., Class C., Whalen W., Jicha B., Bizimis M. and Alvarado G. E.
1062 (2015) Recycled crust in the galápagos plume source at 70 ma: Implications for plume
1063 evolution. *Earth Planet. Sci. Lett.* **425**, 268–277. Available at:
1064 http://dx.doi.org/10.1016/j.epsl.2015.05.036.

1065 Turner S. P., Hawkesworth C., Gallagher K., Stewart K., Peate D. and Mantovani M. (1996) Mantle
1066 plumes, flood basalts, and thermal models for melt generation beneath continents:
1067 Assessment of a conductive heating model and application to the Paraná. *J. Geophys. Res. Solid*

Earth **101**, 11503–11518. Available at: http://doi.wiley.com/10.1029/96JB00430.

Vincent E. A. and Smales A. A. (1956) The determination of palladium and gold in igneous rocks by radioactivation analysis. *Geochim. Cosmochim. Acta* **9**, 154–160.

Webber A. P., Roberts S., Taylor R. N. and Pitcairn I. K. (2013) Golden plumes: Substantial gold enrichment of oceanic crust during ridge-plume interaction. *Geology* **41**, 87–90. Available at: http://pubs.geoscienceworld.org/geology/article/41/1/87/131065/Golden-plumes-Substantial-gold-enrichment-of.

White R. S. and McKenzie D. (1995) Mantle plumes and flood basalts. *J. Geophys. Res. Solid Earth* **100**, 17543–17585. Available at: http://doi.wiley.com/10.1029/95JB01585.

Witten I. H., Frank E. and Hall M. a (2016) *Data Mining: Practical Machine Learning Tools and Techniques*. 4th ed., Elsevier. Available at: http://books.google.com/books?id=bDtLM8CODsQC&pgis=1.

Wittig N., Webb M., Pearson D. G., Dale C. W., Ottley C. J., Hutchison M., Jensen S. M. and Luguet A. (2010) Formation of the North Atlantic Craton: Timing and mechanisms constrained from Re-Os isotope and PGE data of peridotite xenoliths from S.W. Greenland. *Chem. Geol.* **276**, 166–187. Available at: http://dx.doi.org/10.1016/j.chemgeo.2010.06.002.

Zuo R. (2017) Machine Learning of Mineralization-Related Geochemical Anomalies: A Review of Potential Methods. *Nat. Resour. Res.* **26**, 457–464.

Figure Captions -

Figure 1 – a) Schematic map of the North Atlantic Igneous Province. Onshore and offshore lavas are shown in orange, with sample localities highlighted. Pt/Pd for each locality are shown in the purple to light blue bubbles, lightening towards lower values. The orange dotted line describes the trail of the proto-Icelandic plume from c. 65 Ma onwards *(based on Lawver & Muller, 1994)*. Study localities from which data are drawn are shown in white boxes. b) Timeline of eruption in localities in the NAIP throughout the last c. 62 Ma, divided into pre-, syn-, and post-rift periods. Adapted from Hughes et al. (2015).

Figure 2 – Geochemical workflow introduced in this study featuring concurrent a) traditional and b) data science based methods towards a geochemical-geodynamic model. Dimensionality reduction is used to create more concise, descriptive multi-element clusters using MLA.

Figure 3 – Correlation Matrix for all eleven elements included in the bulk geochemical data set for the North Atlantic Igneous Province. Both individual and group correlations are indicated on a scale from -1 to 1, highlighting multi-element trends in a simple manner prior to data science analyses.

Figure 4 – PCA for the NAIP data set. a) Scree plot of eigenvalues for the eleven created PCs and corresponding cumulative percentages; b) Biplot for PC1 v PC2; c) Biplot for PC3 v PC4; and d) Biplot for PC5 v PC6, with variables (elements) plotted as vectors and samples plotted as dots coloured by their respective geographic grouping in all biplots.

Figure 5 – t-SNE plots based on eleven variables, with data points coloured by the relative values of a given constituent element for all eleven analysed variables and Pt/Pd. Dark blue represents the lowest values for a given variable and yellow represents the highest values. Note that Rh and Pt have significant outliers and otherwise less distinctive trends.

Figure 6 – Davies-Bouldin Indices (DBI) for all parameter setups run for k-means clustering models, with the chosen model (*k*=3, using PC1-6 as input variables) highlighted by the black square. By using PC1-6 as variables, the clusters formed are more efficient than using raw data in almost all cases.

Figure 7 – Selection of bivariate PC1 v PC2 plots displaying selected input parameter configurations for *k*-means clustering models. a) *k*=2, PC1-6 as variables; b) *k*=3, PC1-6 as variables; c) *k*=4, PC1-6 as variables; d) *k*=5, PC1-6 as variables. Cluster numbering is randomised by the algorithm and the group order is not relevant to interpretation.

Figure 8 - Bivariate geochemical plots of all NAIP data points, clustered by *k*-means clustering (*k*=3, PC1-6 as variables in the clustering process): a) MgO v TiO_2 b) Fe_2O_3 v Cr c) Ni v Cu d) MgO v Ir e) MgO v Ru f) MgO v Rh

Figure 9 – Bivariate geochemical plots of all NAIP data points, clustered by *k*-means Clustering (k=3, PC1-6 as variables in the clustering process): a) Pd v Pt b) Pd v Pt/Pd.

Figure 10 – Histogram of MLA-based cluster distribution within each geographic category

Figure 11 – Comparison of MLA-based (clustering) groupings and geographic location categories from Hughes et al (2015): a) MLA-based Pd v Pt with average Pt/Pd for each group; b) geographic-based Pd v Pt with average Pt/Pd lines for major trends (with values for W Greenland/BPIP in blue, E Greenland (onshore and offshore) in orange and Iceland in purple); c) MLA-based Ir v Pd/Ir; d) geographic-based Ir v Pd/Ir. Sub-plots (c) and (d) show trends for olivine accumulation and incompatible fractionation in the system, based on (Hughes et al., 2015). Skaergaard (Vincent and Smales, 1956) and Primitive Upper Mantle (PUM) (McDonough and Sun, 1995) estimates are given for reference.

Figure 12 – t-SNE plots with data points assigned to a) *k*-means clustering groups using *k*=3 and PC1-6 as input variables; b) geographic localities. 1 – Tight sub-cluster dominated by Group 1 and East Greenland (off.) (n = 10). 2 – Sub-cluster of Group 3 and East Greenland (off.) (n = 8). 3 – Sub-cluster

of Group 1 and West Greenland (n = 6). 4 – Sub-cluster of Group 3 and West Greenland (n = 13). 5 – Single point defined by Group 1, East Greenland (off.) and high Rh concentration in Figure 5 (n = 1).

Figure 13 – Box-and-whisker plots for all elemental concentrations in each MLA-based cluster (*k*=3, PC1-6 as variables). Displays interquartile ranges (within box), minimum and maximum values excluding outliers (whiskers) and outliers (circles for lower outliers, triangles for extreme outliers) for: a) major oxides; b) trace elements; and c) PGE.

Table Captions -

Table 1 - Summary of data used in this study listed according to geographic regions, localities, reference and number of basaltic lava samples (n).

Table 2 - Summary of elemental concentrations (presented as a range from lower to upper quartile as per Fig. 13, excluding outliers).

Supplementary Items –

Supplementary Data A – Excel sheets for the data set used in this study. The first sheet is the formatted amalgamated data with rounded-zero imputation applied (e.g., no zeroes and non-numeric cells) and calculated z-scores. The second sheet is the original data amalgamated with no processing.

Supplementary Data B – Principal Component Analysis information including eigenvalues, eigenvectors, correlation matrix, and scaled co-ordinates. Generated using ioGas software in conjunction with Python code (Supp. E).

Supplementary Data C – Extra k-means clustering models, displayed in bivariate plots and re-creations of in-manuscript figures. These are models ultimately not selected in the manuscript due to high Davies-Bouldin Indices or inefficient clustering.

Supplementary Data D – Master clustering sheet, showing the classification for a variety of different *k*-means clustering input set-ups. The first sheet shows clusters generated for different *k*-values using z-scored raw elemental concentrations as variables. The second sheet shows clusters generated for different *k*-values using Principal Components as variables. The third sheet shows clusters generated for different *k*-values using t-SNE features as variables (this is ultimately not used in the manuscript as t-SNE cannot be soundly clustered using Euclidean distances and is incorrect but provided as a point of interest).

Supplementary Data E – Jupyter notebooks for Python code for PCA, t-SNE and *k*-means clustering using z-scored raw data and PCs as input variables. Easy-to-follow instructions are written on each for subsequent usage.

Figure 1

Click here to access/download;Figure;Figure_1.jpg

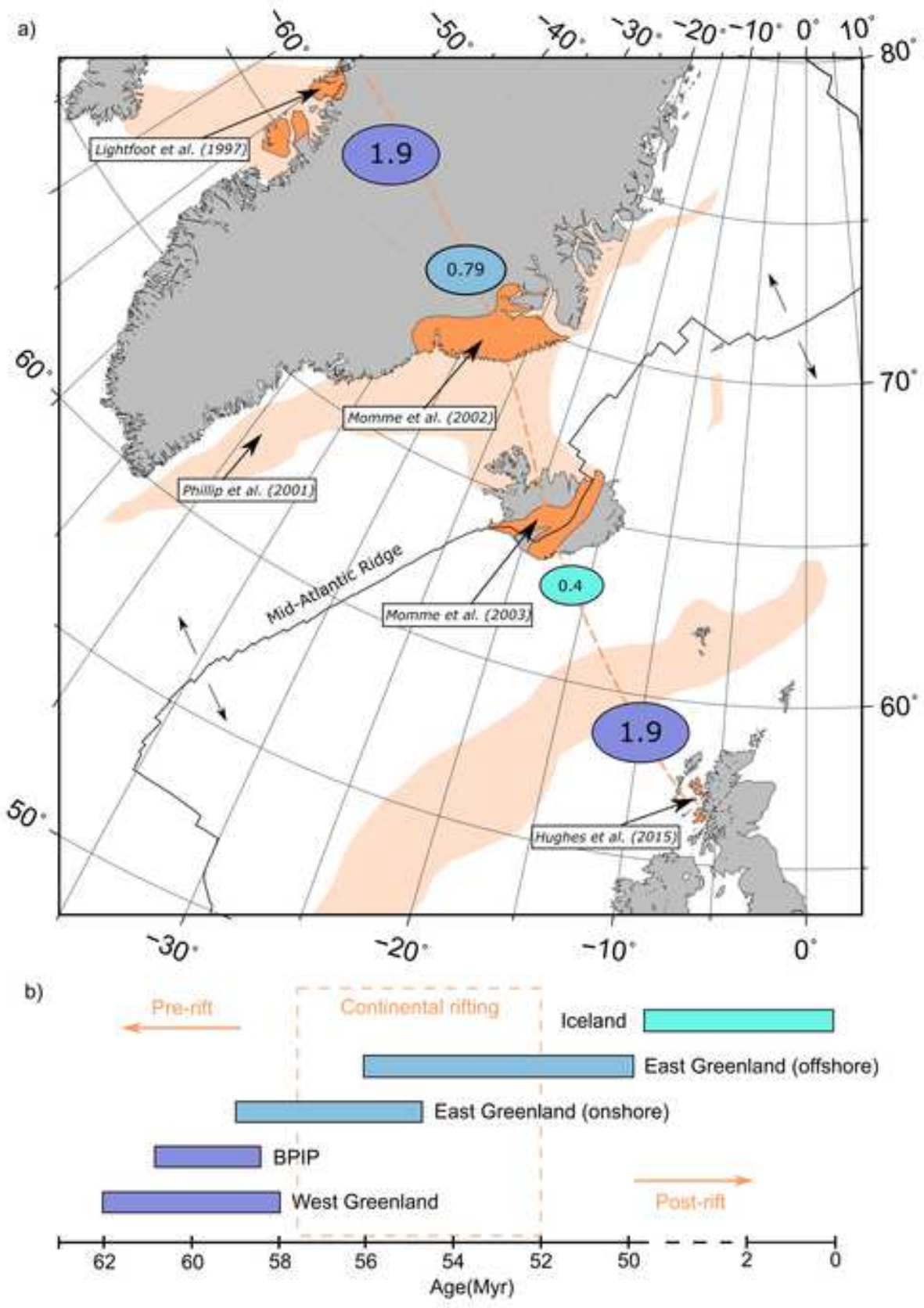


Figure 2

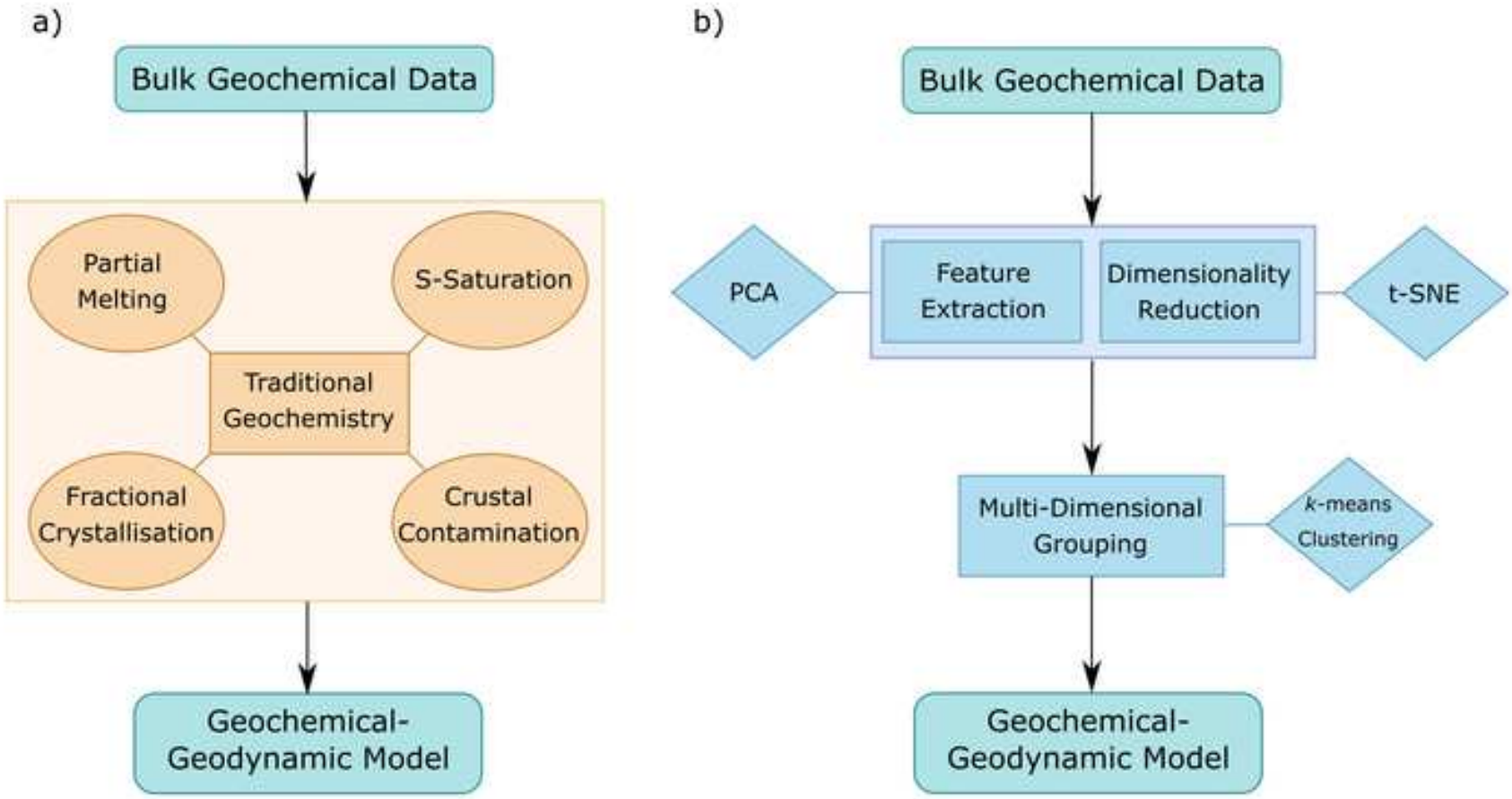


Figure 3

Click here to access/download;Figure;Figure_3.jpg

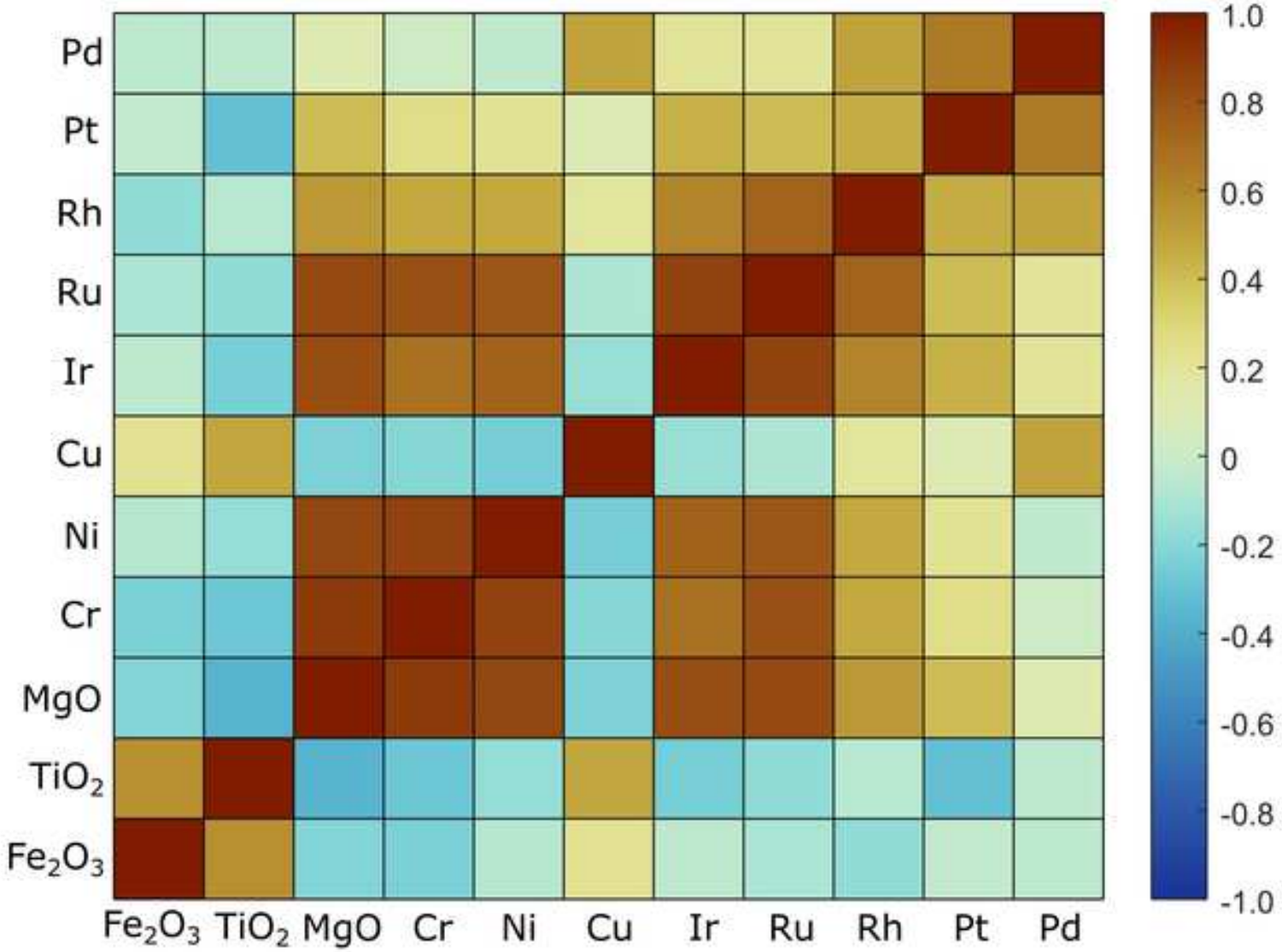


Figure 4

Click here to access/download;Figure;Figure_4.jpg

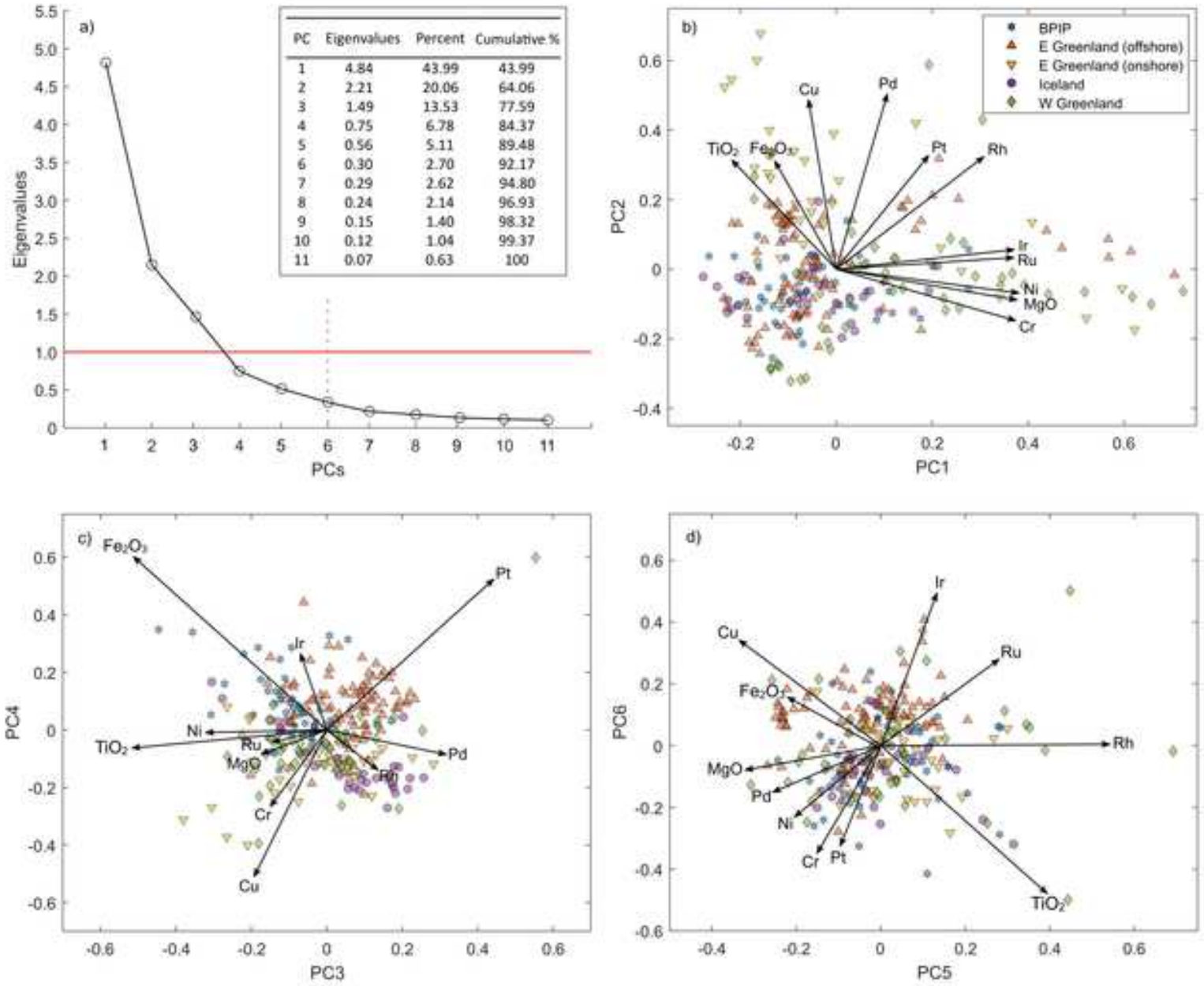


Figure 5

Click here to access/download;Figure;Figure_5.jpg

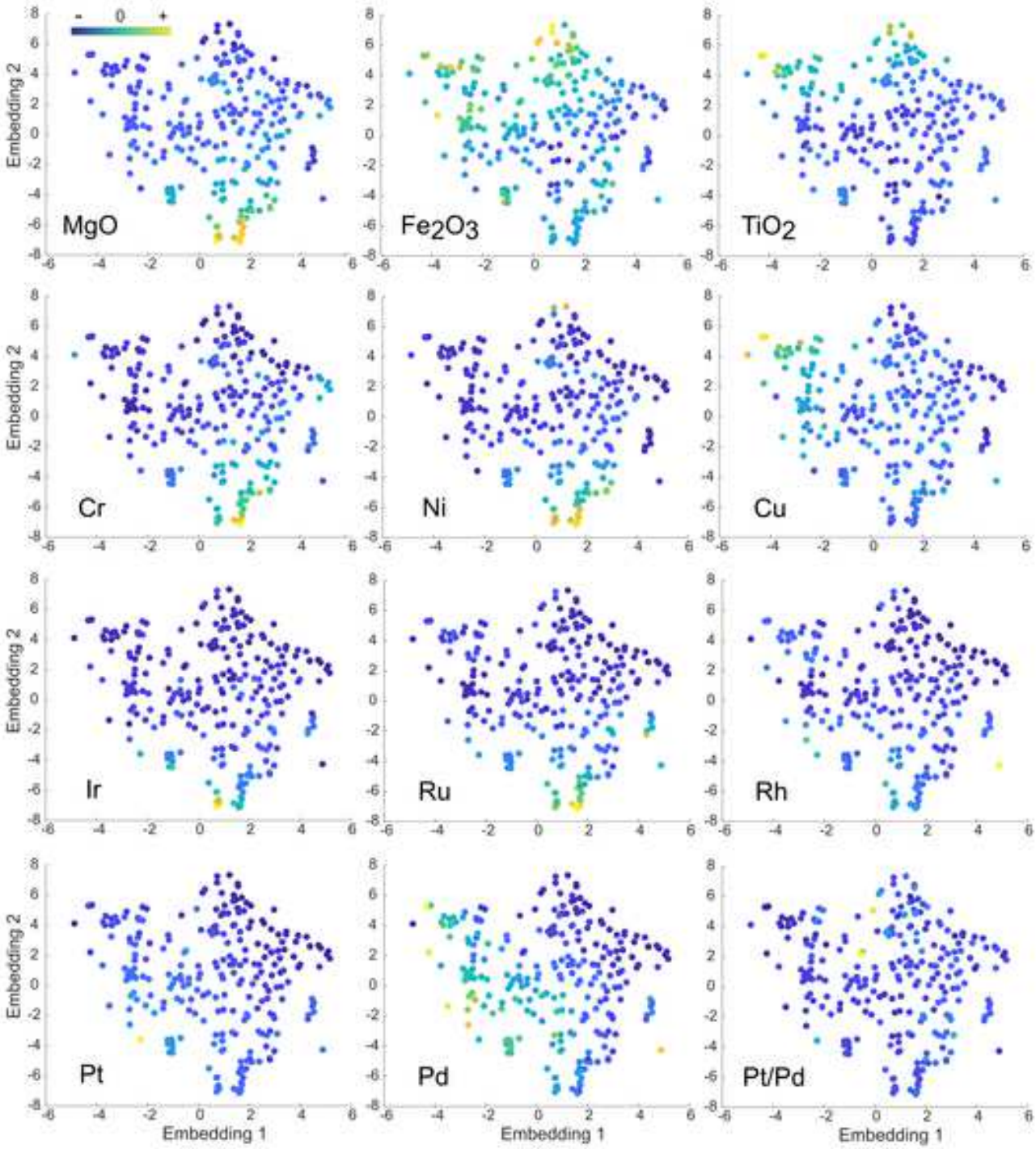


Figure 6

Click here to access/download;Figure;Figure_6.jpg

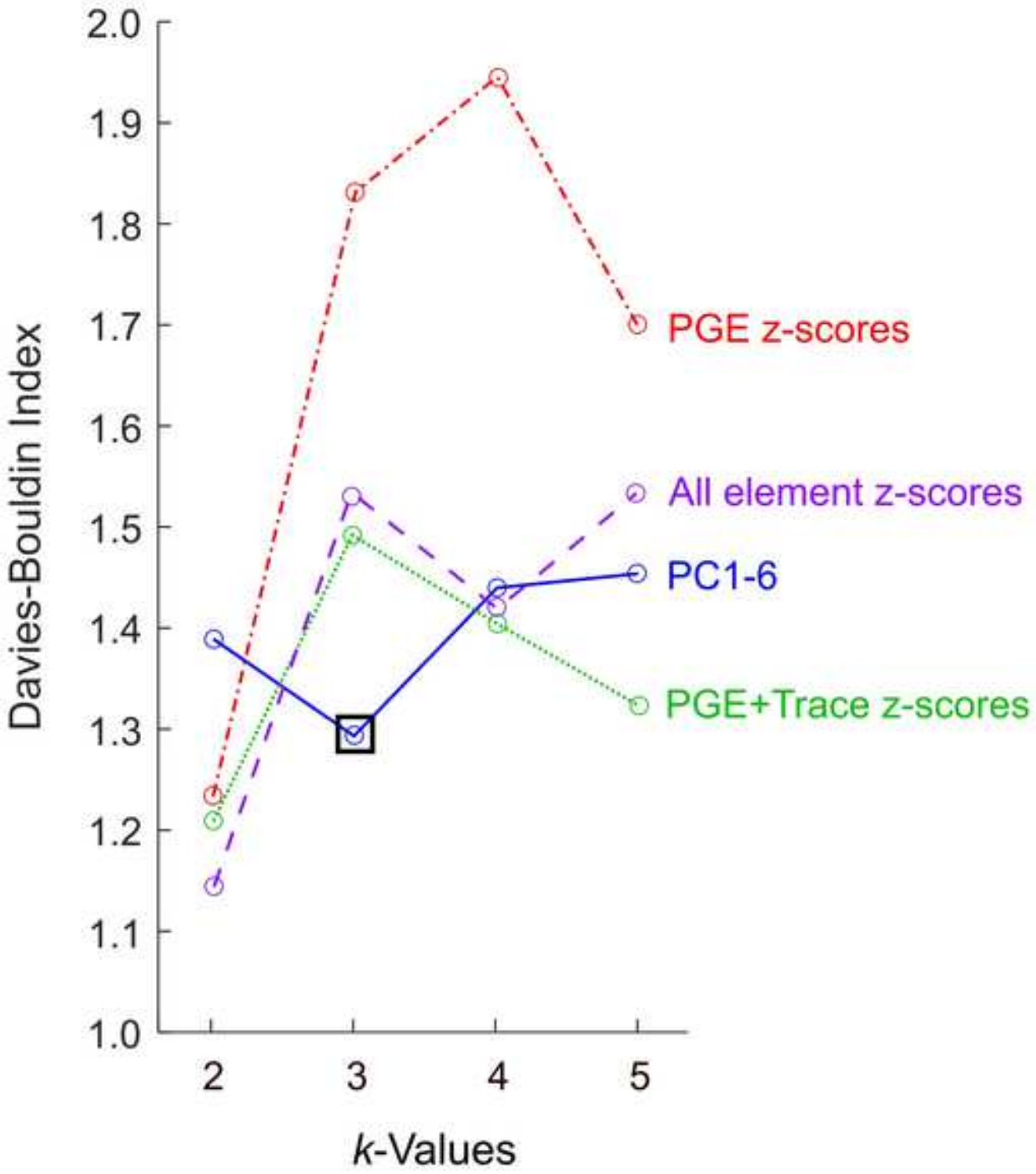


Figure 7

Click here to access/download;Figure;Figure_7.jpg

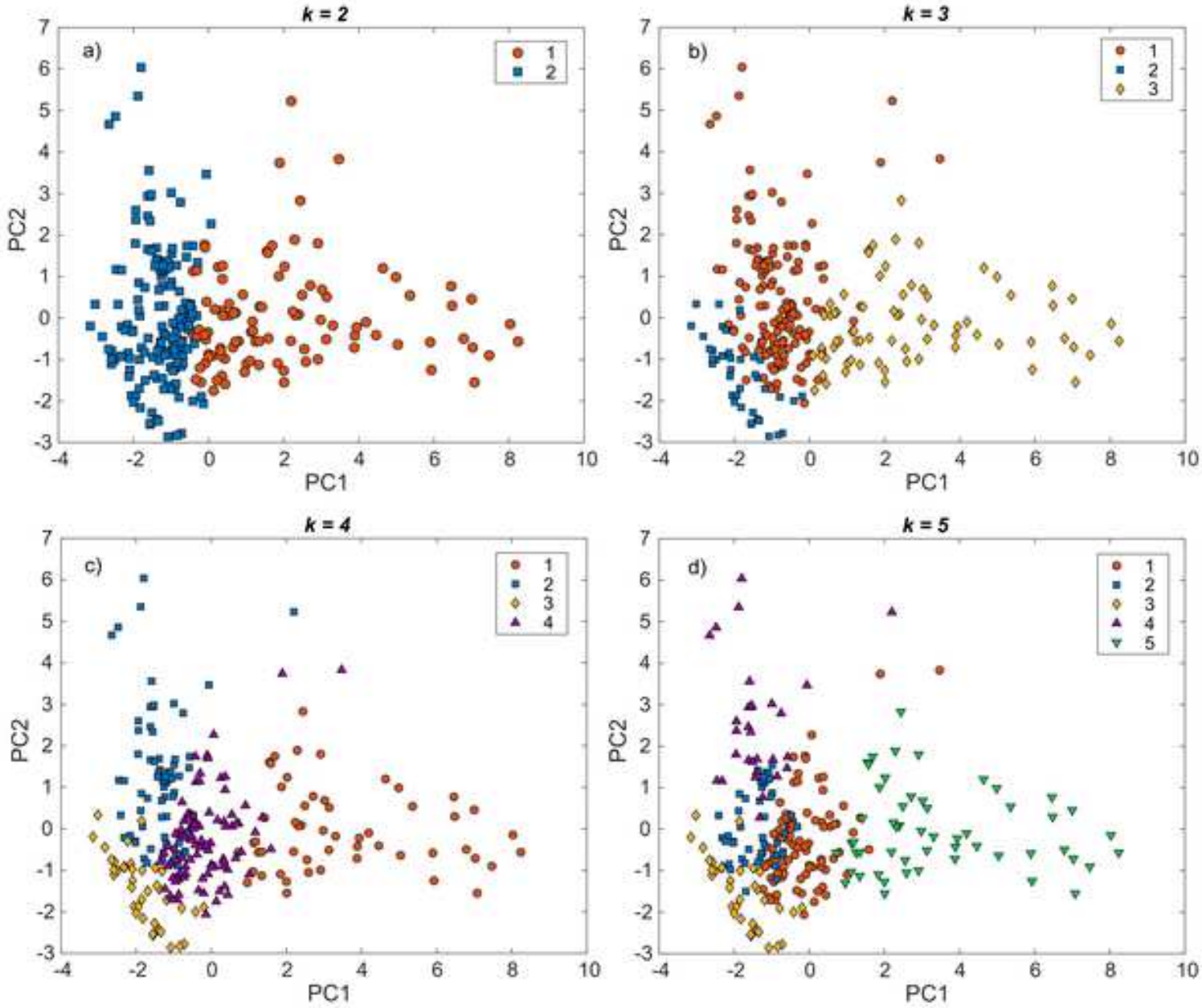


Figure 8

Click here to access/download;Figure;Figure_8.jpg

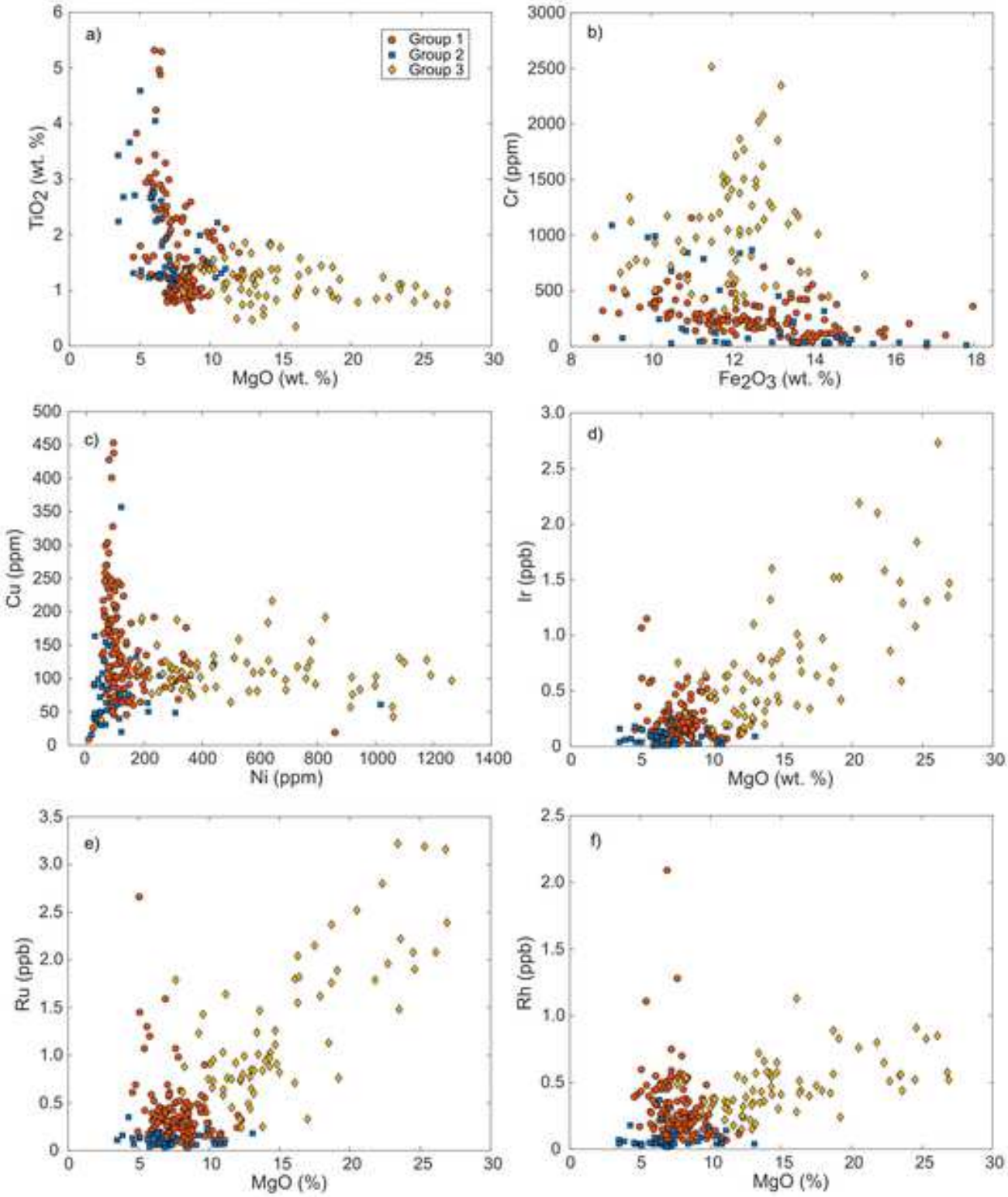
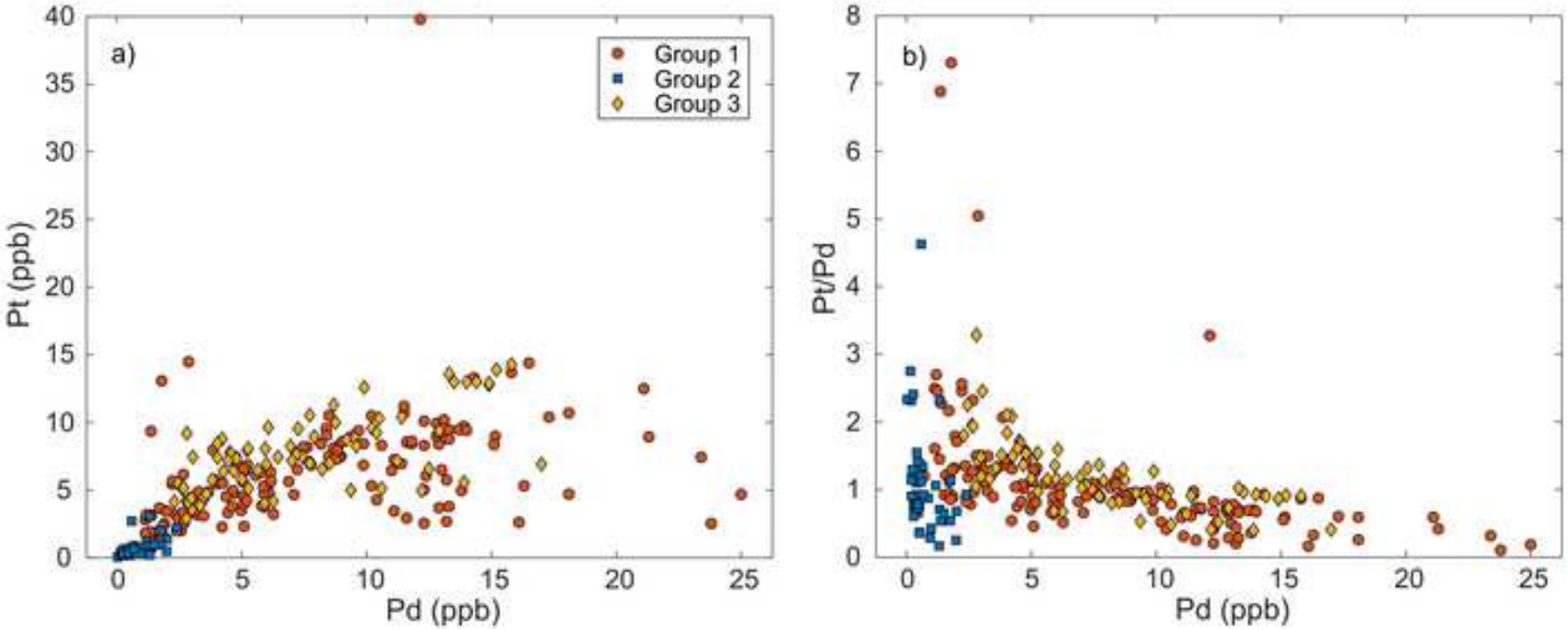


Figure 9

Click here to access/download;Figure;Figure_9.jpg



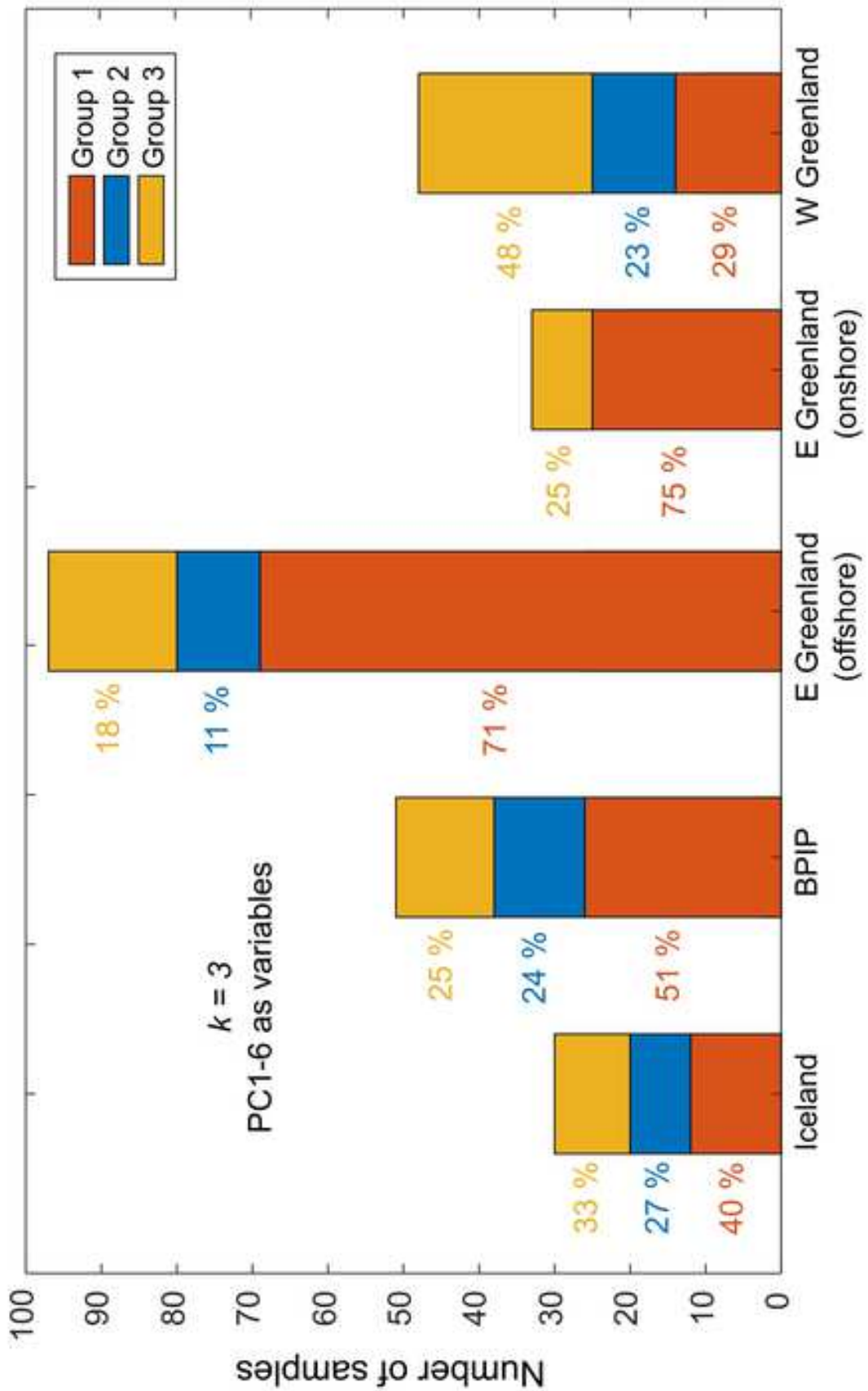


Figure 10

Figure 11

Click here to access/download;Figure;Figure_11.jpg

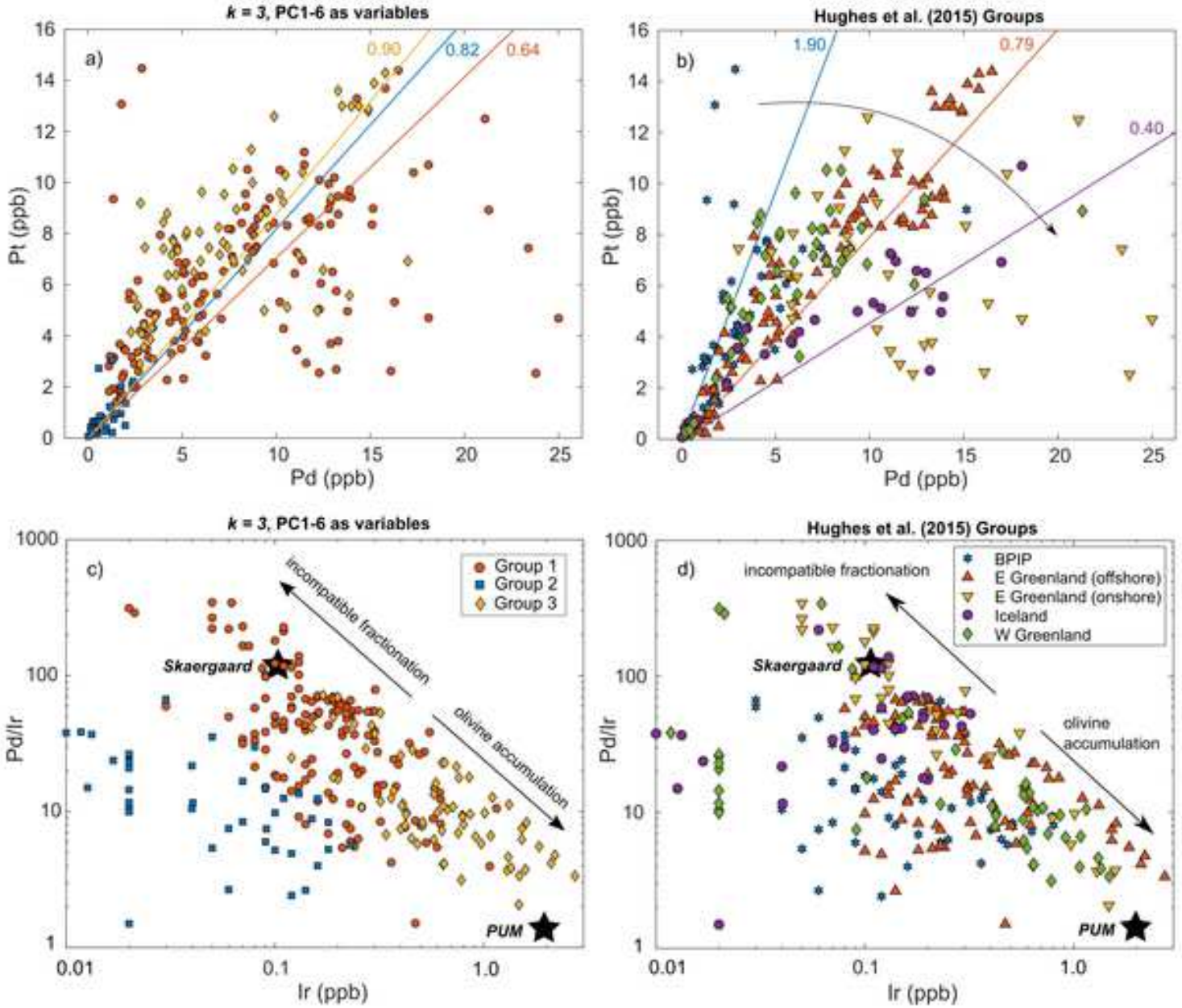


Figure 12

Click here to access/download;Figure;Figure_12.jpg

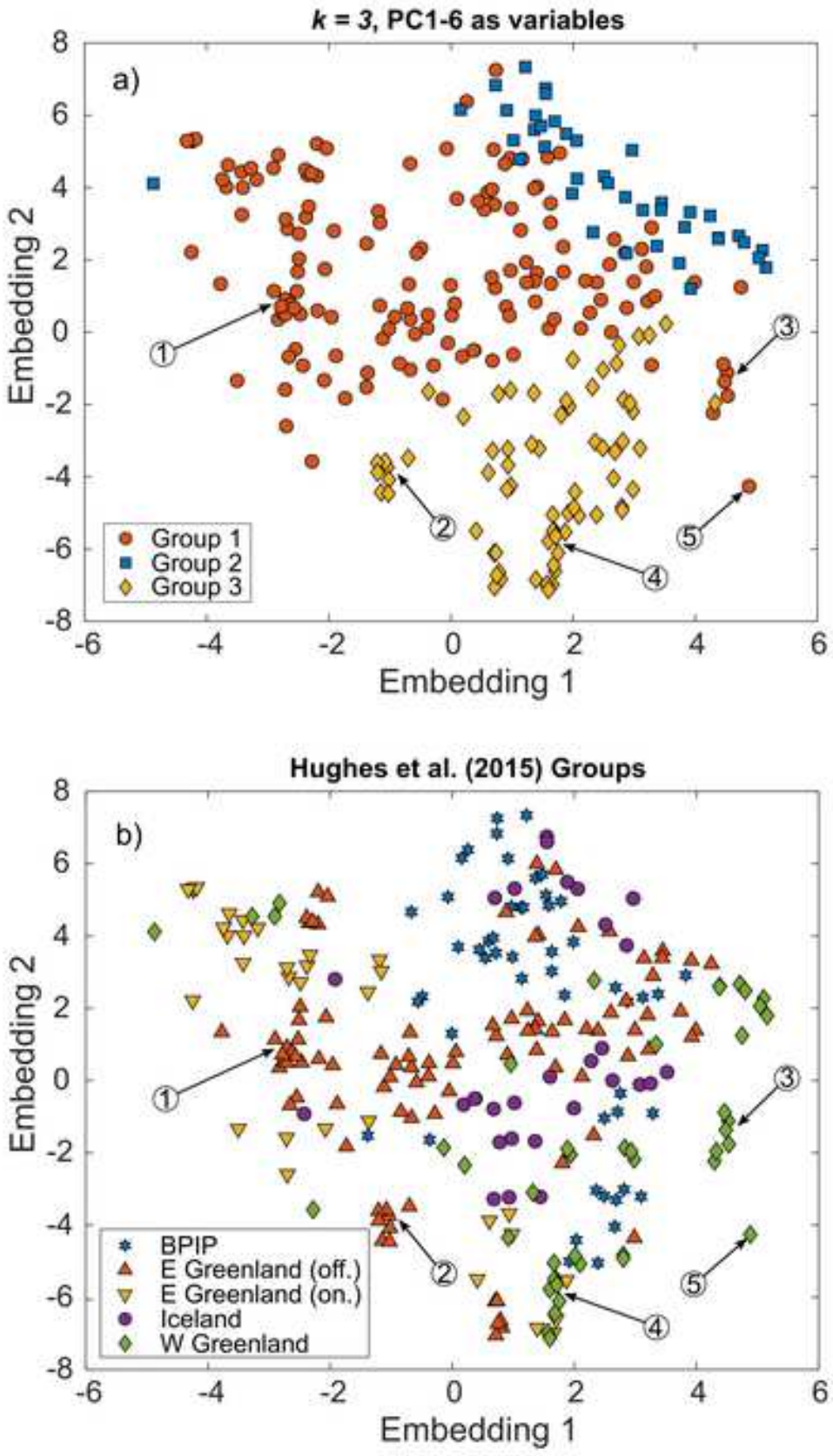


Figure 13

Click here to access/download;Figure;Figure_13.jpg

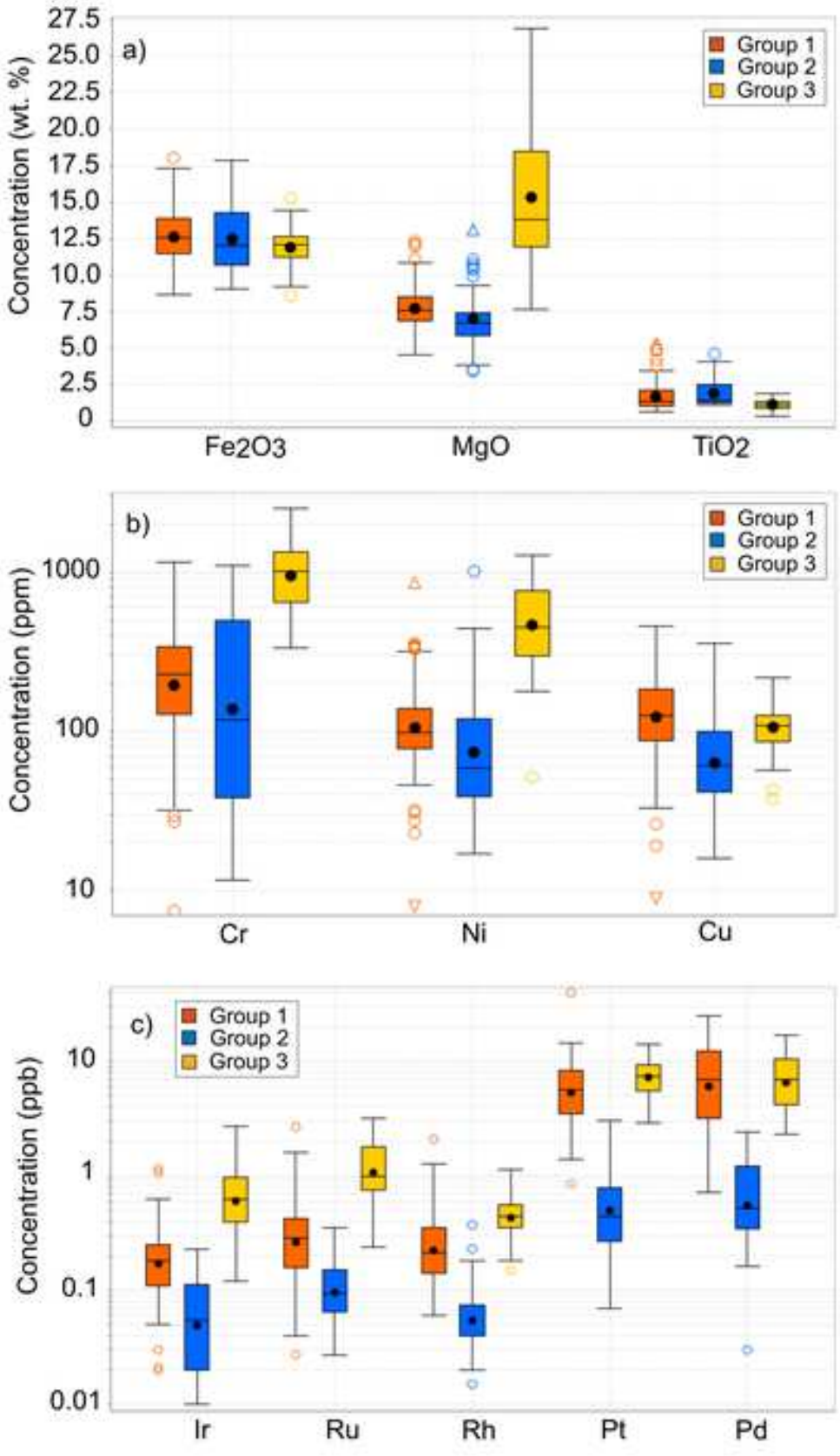


Table 1

NAIP Region	Locality	Study	n
Scotland	Mull, Rum & Skye	(Hughes et al., 2015)	51
West Greenland	Disko Island & Nuussuaq	(Lightfoot et al., 1997)	48
East Greenland (onshore)	Sortebre flood basalts	(Momme et al., 2002)	33
East Greenland (offshore)	ODP 917-918; ODP 988-990	(Philipp et al., 2001)	97
Iceland	West, East and South Rift Zones	(Momme et al., 2003)	30
		Total	***259***

Table 2

Interquartile Ranges	Group 1	Group 2	Group 3
MgO (wt. %)	6.9-8.5	5.9-7.4	11.9-18.5
TiO_2 (wt. %)	1.0-2.1	1.3-2.5	0.9-1.4
Fe_2O_3 (wt. %)	11.4-13.9	10.7-14.3	11.2-12.7
Cr (ppm)	128-337	38-495	646-1,340
Ni (ppm)	78-138	39-120	296-763
Cu (ppm)	87-185	42-100	86-126
Ir (ppb)	0.11-0.25	0.02-0.11	0.40-0.97
Ru (ppb)	0.16-0.42	0.07-0.15	0.75-1.79
Rh (ppb)	0.14-0.35	0.04-0.07	0.35-0.56
Pt (ppb)	3.49-8.40	0.27-0.78	5.51-9.39
Pd (ppb)	3.20-12.19	0.35-1.20	4.22-17.0